Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

Turning Numbers into Knowledge

MASTERING THE ART OF PROBLEM SOLVING

Second Edition

JONATHAN G. KOOMEY, PH.D.

Analytics Press
PO Box 1545
Burlingame, CA 94011-1545
http://www.analyticspress.com
http://www.numbersintoknowledge.com

Analytics Press
PO Box 1545
Burlingame, CA 94011-1545
SAN 253-5602
Internet: http://www.analyticspress.com
email: numbers@numbersintoknowledge.com

© 2008 by Jonathan G. Koomey.
All rights reserved. Protected under the Berne Convention.
2nd edition, 4th printing, September 2013.

Reproduction or translation of any part of this work in any form or by any means, electronic or mechanical, beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the expressed permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to Analytics Press at the address or URL above.

ISBN-13: 9780970601919 (hardcover)
ISBN-13: 9780970601926 (paperback)
Library of Congress Control Number 2008901124

Production manager: Susanna Tadlock and David Peattie
Designer: Sandy Drooker
Compositor: BookMatters
Cover art: Tom Chen
Printer and binder: Thomson-Shore

Text credits:
The graphics characterizing the Cycle of Action in Chapter 3 (Information, intention, and action) are adapted from *The Psychology of Everyday Things* by Donald A. Norman. © 1988 by Donald A. Norman. Reprinted by permission of Basic Books, a member of Perseus Books Group.

This book is printed on acid-free recycled paper (30% post consumer content) in the United States of America.
10 9 8 7 6 5 4

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

FOREWORD
by John P. Holdren

From the late 1970s to the early 1990s, I taught a course originally entitled "Tricks of the Trade" for graduate students in UC Berkeley's interdisciplinary graduate program in Energy and Resources. The course conveyed lessons that I wished someone had taught me during my own university education—but which I mainly learned in "the real world" afterwards—about how to function effectively in a professional life at the intersection of research, analysis, and public affairs.

Berkeley's guardians of academic respectability eventually made me change what they regarded as too frivolous a title for the course to "Professional Methods for Interdisciplinary Careers", but the focus remained the same for the 15+ years that I taught it. It covered ways of thinking through complex problems; how to find and manage information; how to function in a committee; how to identify and avoid common pitfalls in the interpretation of data; how to present results clearly in words, graphs, and tables; how to manage one's time; and even how to avoid jet lag.

Many students over the years suggested that I should write a book teaching the "Tricks of the Trade". Notwithstanding my advice to others about time management, however, I never found the time to write it.

With the 2001 publication of the first edition of Jonathan Koomey's remarkable book, *Turning Numbers into Knowledge*, I realized that I no longer needed to try. Dr. Koomey, who had taken my course in the 1980s as a Berkeley graduate student, had plenty of ideas of his own about the need and how to fill it. And the book that he wrote surpassed what I would have done, if I had found the time, in every important respect.

Now Dr. Koomey has produced a second edition of *Turning Numbers Into Knowledge*, and it is even better than the original. His collection of illustrative

examples, already splendidly germane and instructive, has been expanded and updated, as have his references. His excellent material on the intelligent use of the Web has been augmented with a new chapter on data-sharing websites—their pitfalls as well as their promise. Other useful additions grace nearly every chapter.

In light of all the updates and improvements, even owners of the first edition should want the second. And those who missed the pleasure of the first have the opportunity to start with this still better guide to managing data, time, and people in a problem-solving context. There is nothing else like this book out there. Nobody who deals with problems where numbers matter—and everybody in today's world really needs to—should be without it.

*John P. Holdren**
Woods Hole, MA, October 2007

**Past President, American Association for the Advancement of Science; Director, The Woods Hole Research Center; Teresa and John Heinz Professor of Environmental Policy, Kennedy School of Government, and Professor of Environmental Science and Policy, Department of Earth and Planetary Sciences, Harvard University; and Professor Emeritus of Energy and Resources, University of California, Berkeley.*

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

INTRODUCTION

THE INFORMATION EXPLOSION

In 1815, Thomas Jefferson's library of 6,500 volumes represented an exemplary collection of works by the world's great thinkers. A diligent student who read two books a day, 325 days per year, could peruse the entire collection in ten years. At the time, Jefferson's collection was a large one, and someone who read the library in its entirety could reasonably claim familiarity with the most important knowledge of western civilization.

The same diligent student would have to spend a bit more time to read all the books and manuscripts in the Library of Congress today. At the same rate of 650 books per year, it would take more than 27,000 years to read them all (and that's not even counting the tens of thousands of new books written every year!)[4]

The two centuries since Jefferson, particularly the last half-century, have been extraordinary ones for human knowledge. The pace of change has quickened noticeably, and the rate of information production has increased by many orders of magnitude,[5] driven by rapid progress in science and technology and by vastly improved information transmission, processing, gathering, and storage capabilities.[6] For the first time in human history, information is being created at a rate far faster than humans can assess and use it.

David Shenk, who documents these trends in his book *Data Smog*, reports (among other things) that

> In 1971 the average American was targeted by at least 560 daily advertising messages. Twenty years later that number had risen sixfold, to 3,000 messages per day.
>
> Paper consumption per capita in the United States tripled from 1940 to 1980 (from 200 to 600 pounds) and tripled *again* from 1980 to 1990 (to 1,800 pounds). In the 1980s the use of third-class mail (for sending publications) grew 13 times faster than the population.
>
> Two-thirds of business managers surveyed report tension with col-

leagues, loss of job satisfaction, and strained personal relationships as a result of information overload.[7]

In recent years, rapid growth in email spam has only exacerbated these trends, with more than 90% of emails made up of unwanted spam in 2007.

People usually react to such information overload in one of two ways:

- *They keep reading and never shut down the flow of information* in hopes that something new might emerge to help them solve the problem at hand. In this case, people can be paralyzed by too much information.[8]
- *They shut out new information because it's overwhelming*. People afflicted with this reaction are handicapped by too little information. They rely on old knowledge and ideology, and make bad decisions because they close off the flow.

My goal is to help you plot a middle course between these two extremes, giving you tools and tricks to help you face the onslaught of new information with equanimity. The rest of this book will hone your analytical instincts and prepare you to prosper in this information-glutted world.

Our networks are awash in data. A little of it is information. A smidgen of this shows up as knowledge. Combined with ideas, some of that is actually useful. Mix in experience, context, compassion, discipline, humor, tolerance, and humility, and perhaps knowledge becomes wisdom. —**CLIFFORD STOLL**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

PREFACE

Whatever failures I have known, whatever errors I have committed, whatever follies I have witnessed in private and public life have been the consequence of action without thought. — **BERNARD BARUCH**

Quantitative problem solving is the process by which we take *numbers* and transform them into *knowledge*, using our instincts and experience to fill in when we don't have all the answers. Although the technical aspects of this process are taught at many universities, the *art* of problem solving is rarely discussed and even more rarely written down. This book teaches the intricacies of that art and will help you become a first-rate analyst in your chosen field.

After reading this book, you will be well equipped to make independent judgments about analysis used by others. You will know which key questions to ask, so you need never again be at the mercy of those who traffic in "proof by vigorous assertion."[2] You will also be more effective at conducting and presenting your own analyses, no matter what the topic.

Mastering the art of problem solving takes more than proficiency with basic calculations: it requires understanding how people use information and learning about things as diverse as exploring your ideology, telling good stories, and distinguishing facts from values. To give you a feeling for what to expect, I present an annotated chapter list below.

ANNOTATED CHAPTER LIST

This book contains five major sections, separated into 39 short chapters. Each chapter is compact and self-contained, and each summarizes key lessons I've learned over the years.

INTRODUCTION • THE INFORMATION EXPLOSION: This section briefly describes how analysis can help reduce the information overload that affects us all.

PART I • THINGS TO KNOW: These chapters summarize ideas to keep in mind as you read the rest of the book. More experienced analysts should delve into the ones they find most intriguing and skim the rest.

Chapter 1 • Beginner's Mind: Start fresh and approach any problem like a beginner would and you'll surely see things that others will miss.

Chapter 2 • Don't Be Intimidated: The difference between success and failure often depends on whether you are intimidated. By consciously refusing to be cowed you can stack the odds in your favor.

Chapter 3 • Information, Intention, and Action: This chapter describes how humans respond to events, exploring the connections between what we measure, what we assume, and what we choose to do.

Chapter 4 • Peer Review and Scientific Discovery: Progress in science can be subject to human frailty, just as can any other human endeavor. The end result, however, is something you can count on, in large part because of the peer review process.

PART II • BE PREPARED: A key determinant of your effectiveness is the quality of your preparation. Whether you're building a house or chairing a meeting, preparation for the analysis tasks at hand can turn a potential disaster into a triumph.

Chapter 5 • Explore Your Ideology: Ideology provides a simplified model of the world that reflects your values and experiences and prevents paralysis in the face of the myriad choices you face every day. Make sure you know your own belief system and those of others.

Chapter 6 • Get Organized: Working and living in chaos is like running a marathon with your feet tied together. Get your life in shape and keep it that way.

Chapter 7 • Establish a Filing System: Few mistakes are more maddening than knowing you have seen a relevant article and not being able to find it. By creating a good filing system, you can prevent this annoyance from ever happening again.

Chapter 8 • Build a Toolbox: My analytical toolbox is the set of tricks and techniques that I use to solve particular problems. This chapter describes some key tools to consider for your own.

Chapter 9 • Put Facts at Your Fingertips: Every analysis requires data. Unless you've memorized the encyclopedia, you'll want to keep some key reference sources within easy reach. This chapter describes the ones I find most useful.

Chapter 10 • Value Your Time: If someone is wasting your time, they are stealing your life. Identify your most productive times of day and protect yourself from interruptions during those periods. Unplug the phone. Go to the library. Take control of those times!

PART III • ASSESS THEIR ANALYSIS: When faced with the assertions of others, it's good to know the right questions to ask. These chapters summarize hard-won knowledge about deciphering other people's analyses.

Chapter 11 • The Power of Critical Thinking: Careful critical thinking is at the root of all good analysis. When the steps described in this chapter become second nature, you will have mastered its essence.

Chapter 12 • Numbers Aren't Everything: Not everything that matters can be quantified, so make sure the unmeasurable doesn't fall through the cracks.

Chapter 13 • All Numbers Are Not Created Equal: Numbers and calculations characterizing the physical world are almost always more certain than those describing human behavior. Many analysts wrongly imply that forecasts based on economic data are just as solid as science. They aren't, so be forewarned.

Chapter 14 • Question Authority: This catch phrase of the 1960s is still applicable today. Authority figures can be wrong or biased, so investigate their assertions in the same way that you'd examine those of someone with whom you're not familiar.

Chapter 15 • How Guesses Become Facts: Always remember that "official" statistics are based on calculations that are often poorly documented, incorrectly cited, or otherwise hazardous to your intellectual health.

Chapter 16 • Don't Believe Everything You Read: Maintain a healthy skepticism, even of well-established sources. In this age of instant information transmission, rumor and error seem to propagate even more quickly than truth.

Chapter 17 • Go Back to the Questions: Any time you rely on survey data to make an important decision, refer back to the actual questionnaire upon which the survey data are based; otherwise, you risk misinterpreting the data.

Chapter 18 • Reading Tables and Graphs: First check for internal consistency, then see if the results contradict other facts you know to be true. Search for cognitive dissonance; any discrepancy between the author's results and what you already know will help you investigate further.

Chapter 19 • Distinguish Facts From Values: Don't be fooled by technical people who portray their advice as totally rational, completely objective, and value-free. If they have made a choice, they have also made a value judgment.

Chapter 20 • The Uncertainty Principle and the Mass Media: Just as the observer of a subatomic particle can disturb that particle by the act of observation, the observer of an institution can disturb that institution by observing and reporting on it. Members of the media (and the analysts who inform them) should take responsibility for the power they wield.

PART IV • CREATE YOUR ANALYSIS: Everyone develops his/her own techniques for creating cogent analyses, and in this section I summarize those I've learned. The importance of organization, clear thinking, careful definitions, systematic exposition, scrupulous documentation, and consistent comparisons cannot be overestimated. You'll learn about each of these here.

Chapter 21 • Reflect: Free yourself from interruptions and give yourself time to reflect. Without such time, you'll never achieve your full problem-solving potential.

Chapter 22 • Getting Unstuck: Everyone gets stuck sometimes, but this pitfall need not hobble your efforts if you use the tricks in this chapter.

Chapter 23 • Inquire: When faced with a problem outside your expertise, don't surrender! It's an advantage to be unconstrained by the mental shackles most disciplines place on their practitioners. Some of the most important insights in modern thought came from people who could think "outside the box" (or ignore the box entirely).

Chapter 24 • Be a Detective: Detectives are real-world practitioners of the scientific method. The time-honored techniques of these seasoned problem solvers should be grist for your analytical mill.

Chapter 25 • Create Consistent Comparisons: People often relate best to anecdotes. A consistent comparison is a well-chosen set of anecdotes that illustrates your point in a compelling way. It is a powerful technique and one well worth learning.

Chapter 26 • Tell a Good Story: Scenario analysis is the art of structured storytelling, and it's an essential tool for any good analyst. Most people don't realize that this art is both highly developed and pertinent to many everyday situations.

Chapter 27 • Dig into the Numbers: Don't be shy about delving into the actual numbers even if you're a highly paid executive. You'll learn things you'd never see if someone else crunches the numbers.

Chapter 28 • Make a Model: Models are "laboratories for the imagination," and this chapter explores the subtleties of using them to explain the world around you.

Chapter 29 • Reuse Old Envelopes: You can calculate almost anything using only common knowledge–you just need to learn how to put this knowledge to use, and this chapter (which focuses on back-of-the-envelope calculations) is just the thing to help you do it.

Chapter 30 • Use Forecasts with Care: The future is uncertain but people keep trying to forecast it anyway. Numerous pitfalls await, and without a keen eye for the tricks of this trade, you'll be hard pressed to avoid them.

Chapter 31 • Hear All Sides: In any intellectual dispute, it pays to hear two well-prepared debaters argue their points before drawing any conclusions. Always make such debates fodder for your deliberations and your decisions will benefit.

PART V • SHOW YOUR STUFF: Once you've done good work you'll want to present it effectively to readers or listeners. The chapters in this section give insights into making your results "grab" your audience, designing good tables and figures, and using those tables and figures to convey your key points. The section concludes by exploring effective use of the Internet for publishing your analysis and sharing your data.

Chapter 32 • Know Your Audience: Most analysts forget that other people don't care nearly as much about their results as they themselves do, so know your audience and present compelling information in a form your readers can easily grasp.

Chapter 33 • Document, Document, Document: An astounding number of analysts routinely omit vital data and assumptions from their reports, but you should avoid this pernicious practice. The best analysts document *everything*, giving credit where credit is due, leaving a trail so they can remember, and creating a trail for others to follow. Documentation is also a key step in checking your work, because it forces you to think clearly about your analysis.

Chapter 34 • Let the Tables and Graphs Do the Work: When writing technical reports, create the analysis, tables, and graphs first, then write around them. If the analysis is well thought out, the tables and graphs well designed, and the audience clearly defined, the report should practically write itself.

Chapter 35 • Create Compelling Graphs and Figures: Follow Edward Tufte's rules for graphical excellence, and avoid the most common pitfalls in designing charts and graphs. Your goal should be to give to the reader "the greatest number of ideas in the shortest time with the least ink in the smallest space."[3]

Chapter 36 • Create Good Tables: A well-designed table is a work of art; a sloppy one is worse than useless. Make your tables a resource that your readers will keep as a reference for many years to come.

Chapter 37 • Use Numbers Effectively in Oral Presentations: Even veteran presenters show too many of the wrong numbers. Present only those numbers that support the story you are telling, and focus on that story, NOT on the numbers themselves.

Chapter 38 • Use the Internet: The old ways of publishing are fast being supplanted by web-based approaches. Learn about these new tools and put them to work for you.

Chapter 39 • Share and Share Alike: Some kinds of information are more valuable when shared. Although there's still work to be done, web technology is finally automating the data sharing process, allowing you to capture benefits in standardization, efficiency, and greater analytical insight.

CONCLUSIONS • CREATING THE FUTURE: This chapter gives perspective on why we use analysis in the first place. Understanding the world is a prerequisite for making it better!

EPILOGUE • SOME PARTING THOUGHTS: After the first edition of *Turning Numbers into Knowledge* was published in 2001 I encountered some widely believed but erroneous statistics, and this epilogue recounts the lessons I learned in debunking them.

WHO SHOULD READ THIS BOOK

This book grew out of my experience in training analysts over the past two decades. It is written for beginning problem solvers in business, government, consulting, and research professions, and for students of business and public policy. It is also intended for supervisors of such analysts, professors, and entrepreneurs (who may not consider themselves analysts but who need to create analyses to justify their business plans to potential investors). Finally, it covers many topics that journalists who focus on scientific or business topics will find useful.

HOW TO USE THIS BOOK

There is no need to read the chapters in order. Go straight to those that interest you most, but skim the chapters you skip. You just might see something useful there that you did not expect.

Most chapters have "links" to other chapters, with graphical signposts indicating which chapter or major section to investigate for each link (the relevant chapter number appears inside). These signposts look like the link to Chapter 14 that appears in the right margin opposite this line.

14

All Uniform Resource Locators (URLs) discussed in the book are enclosed in triangular brackets, to set them off from the text. They appear as follows: <http://www.lbl.gov>. The brackets and any punctuation marks that precede or follow them are not part of the URL.

All URLs, as well as many key data files, are available in electronic form at

<http://www.numbersintoknowledge.com>. If you have questions, comments, or suggestions you can post them at this site. I'll gladly evaluate them for inclusion in the next edition. I'm particularly interested in examples of large and public analytical blunders by people who should know better, examples of bad or good tables and graphs, and suggestions for how the book can be improved or expanded.

The endnotes contain references, attributions, and detailed information for the interested reader. The *Further Reading* at the end of the book does not attempt to be comprehensive. Rather, it contains selected sources for each chapter that I regard as most crucial for mastering the material. If I mention a book, I include it in the Further Reading section for the chapter in which I refer to it. At the beginning of the Further Reading section, I also include the list of my very favorite sources on this topic, which (in my opinion) are the "must read" items that all serious problem solvers should have on their shelves.

WHAT'S NEW IN THE SECOND EDITION

In my revisions for this edition I've focused on tightening up the text, revising and improving the examples, updating data where appropriate, and updating and expanding the further reading section. I also added a new chapter on data sharing sites as well as an Epilogue, which describes some of what I've learned since the first edition of *Turning Numbers into Knowledge* was published in 2001. Finally, John Holdren graciously consented to write a new foreword. I hope you enjoy reading this book as much as I enjoyed writing it!

Like all other arts, the science of deduction and analysis is one that can only be acquired by long and patient study.

— **SHERLOCK HOLMES**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CONTENTS

FOREWORD xiii
PREFACE xv
ACKNOWLEDGMENTS xxiii

INTRODUCTION • **THE INFORMATION EXPLOSION** 1

PART I • **THINGS TO KNOW** 3
1 • Beginner's Mind 5
2 • Don't Be Intimidated 7
3 • Information, Intention, and Action 9
4 • Peer Review and Scientific Discovery 22

PART II • **BE PREPARED** 29
5 • Explore Your Ideology 31
6 • Get Organized 34
7 • Establish a Filing System 38
8 • Build a Toolbox 41
9 • Put Facts at Your Fingertips 44
10 • Value Your Time 52

PART III • **ASSESS THEIR ANALYSIS** 57
11 • The Power of Critical Thinking 59
12 • Numbers Aren't Everything 62
13 • All Numbers Are Not Created Equal 65
14 • Question Authority 68
15 • How Guesses Become Facts 73
16 • Don't Believe Everything You Read 76

17 • Go Back to the Questions 81
18 • Reading Tables and Graphs 84
19 • Distinguish Facts from Values 87
20 • The Uncertainty Principle and the Mass Media 90

PART IV • CREATE YOUR ANALYSIS 93
21 • Reflect 95
22 • Get Unstuck 97
23 • Inquire 99
24 • Be a Detective 102
25 • Create Consistent Comparisons 105
26 • Tell a Good Story 107
27 • Dig into the Numbers 111
28 • Make a Model 125
29 • Reuse Old Envelopes 129
30 • Use Forecasts with Care 136
31 • Hear All Sides 143

PART V • SHOW YOUR STUFF 145
32 • Know Your Audience 147
33 • Document, Document, Document 149
34 • Let the Tables and Graphs Do the Work 161
35 • Create Compelling Graphs and Figures 166
36 • Create Good Tables 177
37 • Use Numbers Effectively in Oral Presentations 186
38 • Use the Internet 189
39 • Share and Share Alike 201

CONCLUSION • CREATING THE FUTURE 209

EPILOGUE • SOME PARTING THOUGHTS 213

FURTHER READING 223
NOTES 241
INDEX 245

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

PART

I

THINGS TO KNOW

Before we begin our explorations, there are a few preliminary items to work through that can help increase your problem-solving proficiency. You may need to read these chapters a couple of times, but they are worth mastering.

Don't allow yourself to be intimidated by know-it-alls who thrive on bestowing their knowledge on insecure people. Put cotton in your ears and blinders next to your eyes, and trudge ahead with the confidence that whether or not someone else "knows it all" isn't really relevant; the only thing that's relevant is what YOU know and what YOU do.

— **ROBERT RINGER**

CHAPTER

1

BEGINNER'S MIND

In the course of my life I have often had to eat my words, and I must confess I have always found it a wholesome diet.

— **WINSTON CHURCHILL**

Most people attack a new problem by relying heavily on the tools and skills that are most familiar to them. While this approach can work well for problems that are similar to those previously solved, it often fails, and fails miserably, when a new problem is particularly novel or vexing. In these circumstances, it is best to assume nothing and treat the problem as if you have never seen anything like it before.

In martial arts, this sense of looking freshly at something is known as "Beginner's Mind." Beginners to any art don't know what is important and what is irrelevant, so they try to absorb every detail. Experienced martial artists use their experience as a filter to separate the essential from the irrelevant. When that filter mistakenly screens out something essential, then even seasoned masters can err.

The phrase "Beginner's Mind" is a reminder that experience is a two-edged sword. It eliminates unnecessary detail and allows a focus on what are ostensibly the most important elements of a problem. It can also lead you astray when a new problem is sufficiently outside your experience (Chip and Dan Heath call this problem "the curse of knowledge"). The art is in knowing when experience is not applicable to the problem at hand.

The ideal is to combine the curiosity and the non-judgmental observation of a beginner with the experience of a senior analyst. Together, these skills can be an immensely powerful combination.

A Japanese Zen master received a university professor who came to inquire about Zen. It was obvious to the master from the start of the conversation that the professor was not so much interested in learning about Zen as he was in impressing the master with his own opinions and knowledge. The master listened patiently and finally suggested they have some tea. The master poured his visitor's cup full and then kept on pouring. The professor watched the cup overflowing until he could no longer restrain himself. "The cup is overfull, no more will go in."

"Like this cup", the master said, "you are full of your own opinions and speculations. How can I show you Zen unless you first empty your cup?" — **BRUCE LEE**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

2

DON'T BE INTIMIDATED

Brenda Ueland, in her classic treatise on creative writing (*If You Want to Write*), explicitly advocates thumbing your nose at "know-it-alls, jeerers, critics, doubters," several times a day. These intimidators murder talent and crush creativity, and you shouldn't let them get away with it.

Most of us have experienced people who confidently toss out statistics to bolster their arguments. Although there are those who speak only when they are sure of their facts, there are many more who express their opinions even when their ideas are ill founded. They do so secure in the knowledge that most people will not see through this game.

In 1974, Robert J. Ringer first published his book titled *Winning Through Intimidation*. In it, he described how he developed his own philosophy for success in his chosen field of real estate sales. The core of this philosophy is what Ringer calls "The Theory of Intimidation," which has wide application to every field of human endeavor. In summary, it states that

The rewards a person obtains are inversely proportional to the degree to which that person is intimidated.[9]

In other words, people do worse in work, life, and love when they are intimidated by others.

To avoid being intimidated, Ringer developed a strong posture. He first created a professional image that would prevent others from intimidating him (for example, he created a lavish "calling card" that was so nicely produced that anyone receiving it would immediately be convinced that Ringer was "somebody"). He then perfected the use of legal tools that would ensure his right to claim rewards for his efforts. Finally, he was so good at his profession that no one could legitimately conclude that his performance was anything but excellent.

Ringer's refusal to be intimidated resulted in spectacular financial rewards relatively quickly. The important lessons to take from his book are that

- many people use intimidation to accomplish their goals even when they are not in the right;
- these people often use (and abuse) analysis to commit these acts of intimidation, and
- someone who knows the right question to ask about how such analysis is created can often deflect acts of intimidation before any damage is done.

Part III With the right preparation, you can beat the intimidators at their own game.

Why do people who know the least know it the loudest?

— **GEORGE CARLIN**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

3

INFORMATION, INTENTION, AND ACTION

In spite of people's differences, there are common threads in the way each of us makes decisions and carries them out. The "Cycle of Action" described in this chapter is not an all-encompassing model, but it can help you understand the role of information in your decision making. It can also help you make better decisions in the future.

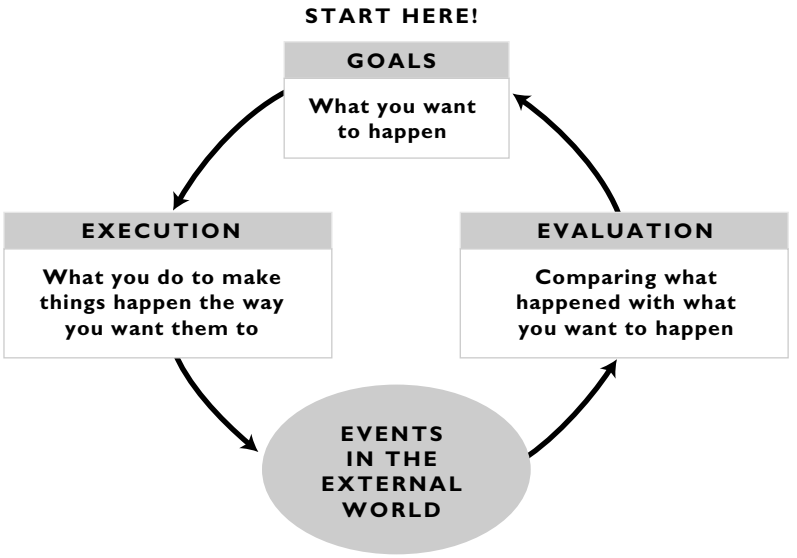
Donald Norman, in his book *The Design of Everyday Things*, first introduced the concept of the Cycle of Action (see **Figure 3.1**) and applied it to how people interact with technology. He separated human responses to events into Goal Setting, Execution, and Evaluation. Say, for example, that you want to hear some music. Goal setting involves deciding which outcome is most desirable (listening to a Beatles tune). Execution is the process of influencing external events so that they move more closely to that outcome (you find the compact disc, put it into the player, and turn the player on). Evaluation is the process of observing the resulting external events and determining whether those events are consonant with your goals (if the music starts and you like what you hear, then you've successfully achieved your goal; if not, you need to chose a new song; if the music doesn't play, you'll need to fix the stereo).

Your decisions are a function both of your past experiences, evaluated after the fact, and your current needs and desires. Your past experience enjoying particular recordings, for example, combined with your current goals of hearing music, will help you decide whether to choose one compact disc over another.

The major steps in the cycle can be summarized as follows:

- Decide what to do, based on your goals;
- Do it (execution);
- Assess the results (evaluation);
- Decide what to do next, and then repeat the process

FIGURE 3.1. The Cycle of Action


START HERE!
GOALS
What you want to happen
EXECUTION
What you do to make things happen the way you want them to
EVALUATION
Comparing what happened with what you want to happen
EVENTS IN THE EXTERNAL WORLD

SOURCE: Adapted from Norman, *The Design of Everyday Things*

Norman applied the cycle to characterize the choices of individuals, but he could just as well have used it to describe the process by which institutions interact with the external world. Of course, no person or institution follows such a methodical procedure all the time, but you'd be surprised at how much such a
28 simple model explains. Let's delve into more detail, and the value of this model will become obvious soon enough.

There's no identifiable beginning to the cycle (except perhaps the day you were born), but, for convenience, we'll start with Goal Setting.

THE GOAL SETTING PROCESS

You set goals by deciding what you want to happen, based on your values, evaluation of past experiences, and assessment of the current situation. By its very nature, this process cannot be totally objective or value free. If it involves
19 choices, it involves value judgments. To pretend otherwise is both foolhardy and shortsighted.

Peter Hess and Julie Siciliano, in their book *Management: Responsibility for*

Performance, define key characteristics of effective goal setting in the management context, but these criteria apply more generally. The goals must be specific, meaningful to and accepted by the participants, and "realistic yet challenging." Finally, effective goals have a clearly defined deadline and ways of measuring progress toward the goal.[10]

The goal can be as simple as satisfying your hunger or as complicated as influencing the future direction of the computer industry. Once you've decided what outcome you desire, you take action (execute).

THE EXECUTION PROCESS IN DETAIL

The execution part of this cycle can be split into three pieces, as shown in **Figure 3.2**:

- your intention and desire to achieve the goals, which exists because of your human need for closure, fulfillment, status, etc.;
- a list of actions (the "action sequence") that you plan to undertake, which is developed based on your evaluation of the situation, your goals, and your experience; and
- the actual execution of the action sequence, successful completion of which depends on your capabilities.

Developing an intention to act

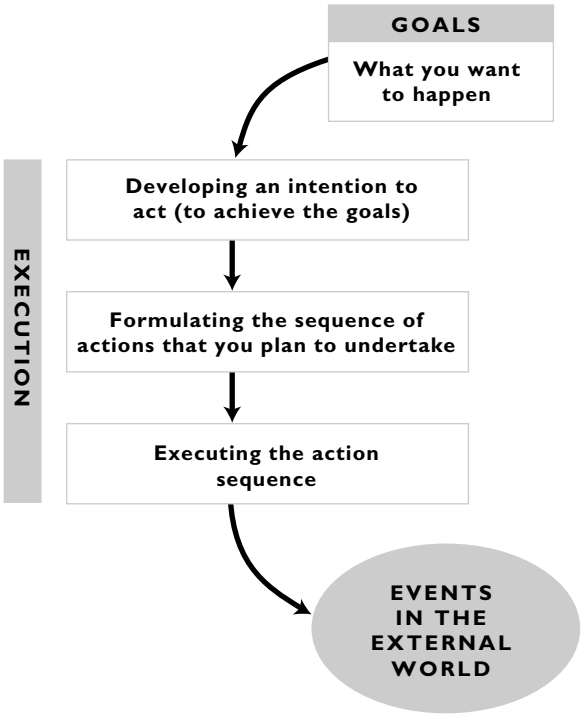
Execution cannot take place without an intention to act. Before deciding what to do, you first need to decide to do *something*. From this intention flows the next step (a decision about *which* action to take). If you decide not to act, the cycle stops here.

Formulating the action sequence

The action sequence is the set of steps that, if taken in the correct order, will lead to the desired outcome, assuming no unanticipated issues arise. The decision maker develops this sequence based on judgment, experience, and forecasts of likely alternative outcomes for different courses of action.

Accurate data collection and analysis make it more likely that the action

FIGURE 3.2. The execution part of the cycle of action


GOALS
What you want to happen
EXECUTION
Developing an intention to act (to achieve the goals)
Formulating the sequence of actions that you plan to undertake
Executing the action sequence
EVENTS IN THE EXTERNAL WORLD

SOURCE: Adapted from Norman, *The Design of Everyday Things*

sequence will be developed in a sensible way in any particular instance. Only rarely does the decision maker have complete knowledge of all relevant factors, so she relies on instinct, experience, ideology, and whatever relevant data have been brought to bear on the topic. Any decision is usually one of dozens to be made on a given day, and there are few such decisions that do not also face time constraints. Poor judgment can result as much from tight time constraints as from data that are compiled inaccurately.

Because all information is uncertain in one way or another, it is important to develop strategies to cope with that uncertainty when formulating the action sequence (for more details, see Morgan and Henrion's book *Uncertainty*). One of the most effective coping strategies involves thinking through alternative scenarios in advance of a particular decision.

When evaluating how information was used to develop the action sequence in a particular instance, always explore the interests and training of the people involved. Such an exploration will almost certainly be fruitful. In his book *How to Drive Your Competition Crazy*, Guy Kawasaki points out that "a manager who has been through years of price wars will interpret an opponent's price reduction differently from a Harvard MBA, who's never been closer to a price war than a lecture hall, or a Wharton MBA, who has to formulate an econometric model to buy a pack of gum."[11]

Executing the action sequence

After deciding what to do, you must follow through with action. Good data documentation practice, accurate analysis, and critical thinking can help ensure that the outcome of execution is what you intend. The perceptions of others can be key to success, and a decision maker backed up by credible analysis can influence those perceptions to her advantage.

The execution process itself can also affect goal setting. For example, the very process of taking action can cause you to realize more precisely what you want. It will also give you better information on what it will take to achieve your goal and help you decide whether it's worth continuing to pursue.

EFFECTS ON THE EXTERNAL WORLD

After the action sequence is executed, the decisions affect the external world in some way. For the Cycle of Action, the relevant external world is everything external to the decision process of the individual or institution that might be affected by the action sequence. These effects can be confined to physical ones (i.e., you turn on the stereo after determining the correct action sequence to do it) or to personal/emotional ones (e.g., you decide to play a song on the stereo that reminds you of a long-lost love and brings back your feelings for that person). Usually, though, the effects are both physical and emotional. Often there are many distinct effects of a given action sequence, some of which cannot easily be foreseen. You assess these effects in the evaluation stage of the Cycle of Action.

THE EVALUATION PROCESS IN DETAIL

The Evaluation part of this cycle also can be split into three components, which are illustrated in **Figure 3.3**:

- perceiving the state of the external world, which involves observation, direct measurement, or indirect data collection;
- interpreting the perception, which requires that you apply your knowledge and logic to understanding what you have perceived; and
- evaluating the interpretation and comparing it with your goals. This latter process is the one where value judgments and assumptions most forcefully exert their influence.

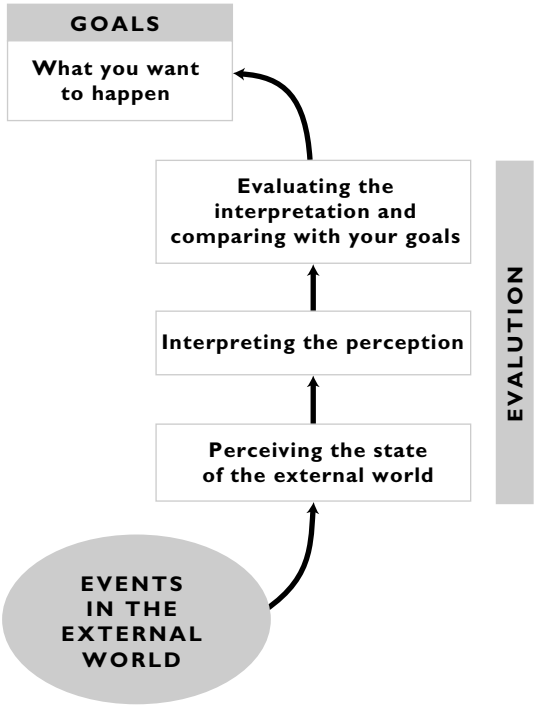
Each step in the evaluation process can be influenced by personal biases, which affect the things you choose to measure and how you interpret those measurements. Preconceptions inhibit data collection and cause unusual results to be
1 dismissed too easily. Remember to approach evaluation using "Beginner's Mind," and you'll fall prey to this pitfall less often.

Perceiving the state of the external world

12 Some aspects of the external world can be measured, but others cannot. Perceiving the external world may include both intuitive and quantifiable knowledge, and each person usually draws on both kinds of knowledge when assessing the current state of the world.

People often measure only those things that are easiest to measure. Like the man looking for his keys under the street lamp (because that's where the light is) even though he lost his key chain down the street, you are likely to be led astray if what you're measuring is only peripherally related to what you care about. If the people charged with collecting the data are too far removed from the actual decision and don't know how the data will ultimately be used, the results can be bad or even disastrous. Guy Kawasaki reports that Sam Walton, the founder of WalMart, did his own research on the competition as he was building his business because *he* was the one who ultimately had to make the strategic decisions. He didn't want data collection choices made by people in his organization who did not understand the big picture.[12]

FIGURE 3.3. The evaluation part of the cycle of action


GOALS
What you want to happen
Evaluating the interpretation and comparing with your goals
Interpreting the perception
Perceiving the state of the external world
EVALUTION
EVENTS IN THE EXTERNAL WORLD

SOURCE: Adapted from Norman, *The Design of Everyday Things*

Kawasaki also describes how businesses can sometimes deliberately skew their competitors' perceptions of the external world. In the late 1960s, Wilson Harrel used his knowledge of Proctor & Gamble's (P&G's) test marketing plans (which were public knowledge) first to make P&G believe that its new cleaning product would be a big success. He temporarily withdrew his own competing product (Formula 409) from Denver, the test market city, which made P&G's test product sell well. P&G then decided to roll out the product nationwide, based on the "success" in Denver. To thwart this launch, Harrell promoted an extremely low price for a large amount of Formula 409 just as P&G launched its national campaign. Harrel's promotion effectively removed a large number of consumers from the national market for cleaning fluid, because the special price applied to enough cleaning fluid to last a typical household for half a year. P&G discontinued its product less than a year later, because of disappointing sales.[13]

Information technology has made it easier to collect data on events we could only dream about tracking twenty years ago. In 1997, Peter Coffee, a columnist at *PC Week* magazine, advocated that businesses collect data on sales of products by day of the week and even by time of day.[14] The benefits of such tracking include improved customer service, inventory management, cash flow tracking, and staff allocation, but it has only been during the past decade that such data collection has become possible on a large scale.

Even with new technology, however, collecting the relevant data is not always straightforward. The ranking of different Internet web sites in terms of numbers of visitors is critical for web advertisers and other businesses, but small differences in measurement methodology can lead to large differences in these rankings. Internet companies are still struggling to define appropriate measurement methodologies, and these difficulties are unlikely to be resolved soon.[15]

Inside an institution, the process of perceiving the state of the external world can be subverted by the complexities of office politics. The busiest decision makers (CEOs, Presidents, Cabinet Secretaries) are often protected from new ideas by their legions of "handlers," from their Chiefs of Staff to their secretaries. Robert Reich discovered this truth when he became Secretary of Labor in 1993:

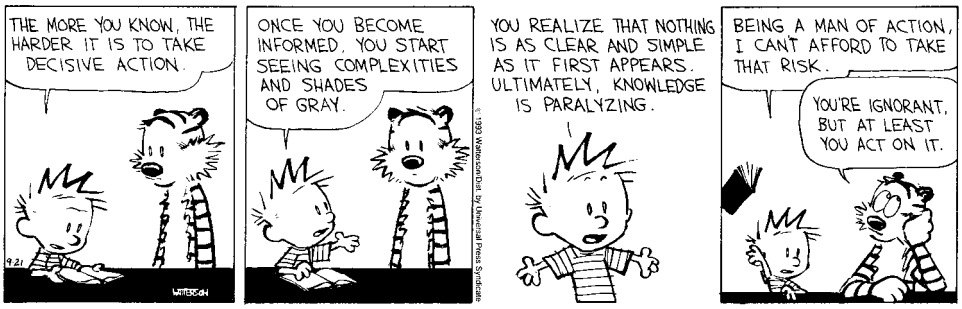
> My cavernous office is becoming one of those hermetically sealed, germ-free bubbles they place around children born with immune deficiencies. Whatever gets through to me is carefully sanitized. Telephone calls are prescreened, letters are filtered, memos are reviewed. Those that don't get through are diverted elsewhere.[16]

The Cycle of Action can thus become the Cycle of Inaction when information is diverted from its intended audience. Never forget that, within institutions, the most important decision makers virtually always face some variant of Reich's "bubble." Reich himself tried to puncture the bubble now and then, to the consternation of his staff.

Interpreting the perception

Once data are collected, they almost always require "massaging" into a useful form. This step can involve as little as minor reformatting or as much as detailed calculations and analysis. The choice of how much to manipulate data

One approach...


THE MORE YOU KNOW, THE HARDER IT IS TO TAKE DECISIVE ACTION.
ONCE YOU BECOME INFORMED, YOU START SEEING COMPLEXITIES AND SHADES OF GRAY.
YOU REALIZE THAT NOTHING IS AS CLEAR AND SIMPLE AS IT FIRST APPEARS. ULTIMATELY, KNOWLEDGE IS PARALYZING.
BEING A MAN OF ACTION, I CAN'T AFFORD TO TAKE THAT RISK.
YOU'RE IGNORANT, BUT AT LEAST YOU ACT ON IT.
WATTERSON

CALVIN AND HOBBES © 1993 Watterson. Dist. by UNIVERSAL PRESS SYNDICATE. Reprinted with permission. All rights reserved.

is as much art as science, and it relies heavily on the instincts of the researcher. This choice is influenced strongly by how particular calculations are to be applied. For analysts who are preparing data for others to use, assessing the needs of those who will use the data is a key step in the data development process. **32**

For the "sales by time of day" example described above, sales data might be viewed for individual products or bundled into groups of products to help discern patterns and trends. The data could also be viewed by hour, by day of the week, by month, or by year. Each view might yield different insights.

If the pending decisions are to be made by people other than those collecting and massaging the data (as is often the case), the data must then be summarized in the form of text, tables, and graphs for the decision makers' use (if the information is for your own use, you can make the tables and graphs but skip the text). This step also involves judgments about which data to emphasize. Poor choices here can make irrelevant the most useful data and can lead to erroneous conclusions. Hourly sales data, for example, might be overwhelming if presented en masse–some filtering is required to make these data useful. Such filtering might include focusing on results for one particularly important product or time period, for example.

As Edward Tufte says, bad table and graph design reflects sloppy thinking. At this stage in the cycle, follow the guidelines presented later in this book about designing figures and tables. Don't complicate the process of "interpreting the perception" by creating tables and graphs that obscure the data and confuse your audience. **35** **36**

Evaluating the interpretation

The evaluation of data involves drawing conclusions, developing assertions, and deciding which of these assertions to emphasize. These assertions can take several forms, but the choice of which ones to make from a given set of data is at the discretion of the analyst (that means YOU!) or the people presenting information to the decision maker. Even among the set of assertions that can be logically supported by a particular set of data, there is no single correct choice here.

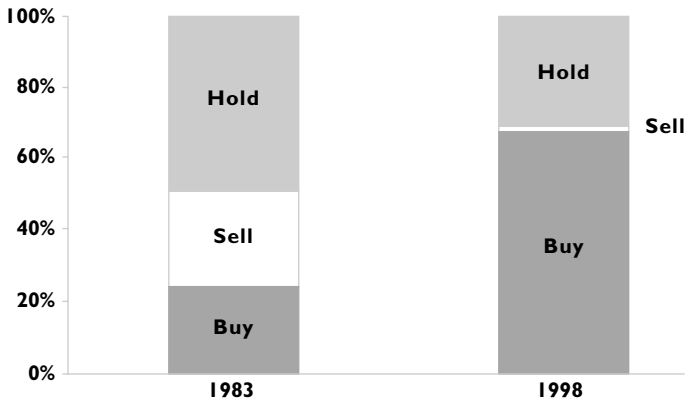
At this point in the cycle, the analyst must make explicit assumptions about
5 19 relevant unknowns (based on experience, values, instinct, and ideology), or simply ignore them. Usually, these issues are noted and put aside for future work. Ignoring these unknowns can lead to serious analytical errors, as can making bad assumptions about them. Unfortunately, there's no way to avoid this risk,
26 but you can insulate yourself from it by creating credible scenarios.

The evaluation at this stage can be affected by power relationships between the analyst and the source of the data. Jeffrey Laderman, writing in *Business Week* in 1998, describes how ratings of stock analysts on particular companies changed over the previous fifteen years. **Figure 3.4** summarizes his results. Although "Sell" recommendations were tendered roughly one-quarter of the time in 1983, such recommendations had virtually disappeared by 1998. The percentage of "Buy" recommendations grew from about one-quarter in 1983 to two-thirds by 1998. Laderman writes that "by the end of the 1980s, Buys outnumbered Sells four to one, and by the early 1990s, eight to one."

In part, this shift was the result of the ongoing bull market in the U.S., but it was also partly caused by the increasingly cozy relationships between investment banks and the companies their stock analysts covered. Banks whose analysts dispensed "Sell" recommendations were shut out of competition for lucrative public stock offerings by those companies. Their analysts were also excluded from company briefings (a tactic similar to that used by many American presidents in dealing with reporters over the past five decades). Company pressure undeniably affected (and probably continues to affect) stock analysts' evaluation of the interpretation when assessing company stocks, although some structural changes in these companies have improved the situation.

Once the evaluation stage is completed (you've compared the outcome of execution with your goals), then the cycle begins anew. If things didn't work out as you had hoped, you should start formulating an action sequence that might lead to the desired result next time.

FIGURE 3.4. Percentage of "buy," "sell," and "hold" stock recommendations by U.S. analysts in 1983 and 1998



SOURCE: Laderman, Jeffrey M. 1998. "Wall Street's Spin Game: Stock Analysts Often Have a Hidden Agenda." In *Business Week*. October 5. pp. 148-156.

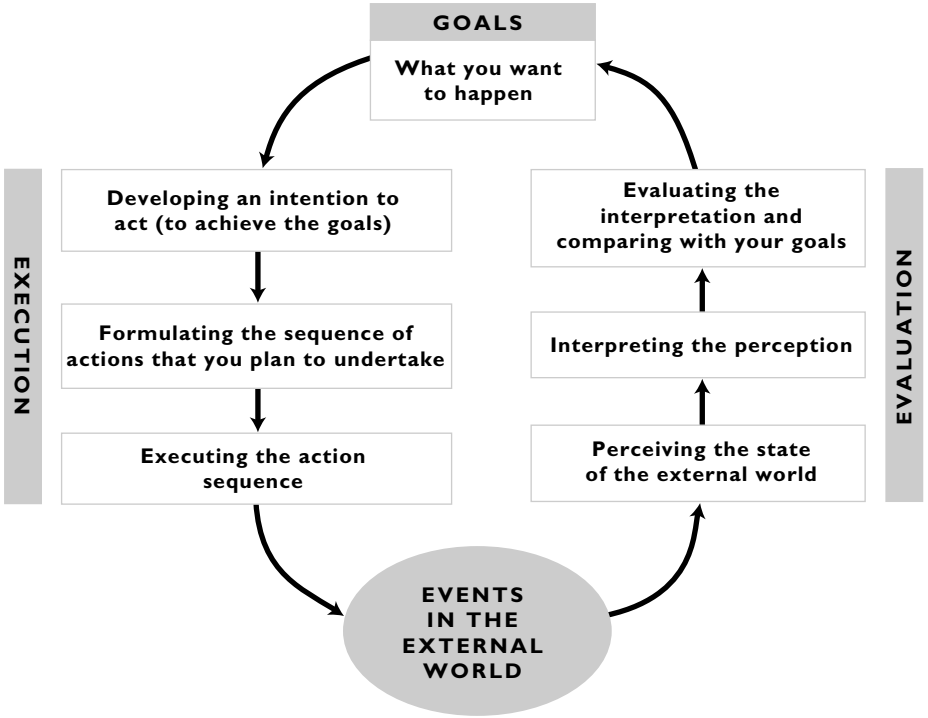
THE BOTTOM LINE

The Cycle of Action reminds us that making decisions and taking action rely on choices about which data to ignore and which to address explicitly. It also illustrates that unknown and unknowable data, as well as experience, intuition, and assumptions, affect *every* choice. No decision is immune from these facts, and you would do well to keep them in mind.

Time constraints can influence all the steps in the Cycle of Action. Rarely do decision makers have enough time to evaluate all the options, so they rely on judgment and ideology to help them assess a situation. Execution of plans also occurs in a time-constrained environment because other people and institutions are making changes that can negate the beneficial effects of your own changes, creating pressure to act quickly.

Refer to the Cycle of Action whenever your choices make events go awry. It can be applied to most decisions in your life, from those about what clothing to wear to those about which college to attend. Systematic assessment of how you use information to make decisions will help you do better in the future.

FIGURE 3.5. The complete cycle of action in detail


GOALS
What you want to happen
Developing an intention to act (to achieve the goals)
Formulating the sequence of actions that you plan to undertake
Executing the action sequence
EXECUTION
EVENTS IN THE EXTERNAL WORLD
Perceiving the state of the external world
Interpreting the perception
Evaluating the interpretation and comparing with your goals
EVALUATION

SOURCE: Adapted from Norman, *The Design of Everyday Things*

CONCLUSIONS

The collection and use of data are human endeavors and, as such, are subject to all the vagaries of individual and institutional behavior. At any point in the Cycle of Action, human frailties can change the course of events. This realization should not inspire fear or resignation. Rather, you should refer back to the complete Cycle of Action in **Figure 3.5** whenever you want to improve your use of information or to explain how a decision-making process broke down in a particular instance. Identify which parts of this cycle were the result of your own choices, those of others, or external events. Such a systematic assessment can help you use information more skillfully and make better decisions in the future.

EXERCISE

Think about three decisions you made in the past week. Did your decision process follow the Cycle of Action in all cases? Where did your biases affect the outcome? How did you decide what your goals were? Did you have enough time to interpret and evaluate your perceptions of the external world before making the decision?

It is not the critic who counts, not the man who points out how the strong man stumbled, or where the doer of deeds could have done them better. The credit belongs to the man who is actually in the arena; whose face is marred by dust and sweat and blood; who strives valiantly; who errs and comes short again and again; who knows the great enthusiasms, the great devotions, and spends himself in a worthy cause; who, at the best, knows in the end the triumph of high achievement; and who, at worst, if he fails, at least fails while daring greatly, so that his place shall never be with those cold and timid souls who know neither victory nor defeat.

— **THEODORE ROOSEVELT**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

4

PEER REVIEW AND SCIENTIFIC DISCOVERY

Science is a human process that advances in fits and starts. The end result of this process is generally something you can count on, but the intermediate steps to this result do not advance in a linear progression as science textbooks might lead you to believe (see Kuhn's *Structure of Scientific Revolutions*). Understanding the process by which scientific knowledge is created can help you better assess its implications and relevance to your own analysis.

Most people (especially nonscientists) don't recognize the large role that intuition and instinct play in scientific discovery, particularly in the process of developing hypotheses. A scientist's desires for recognition and status, enthusiasm for solving a problem, and attitudes towards authority are all motivations rooted in human needs and emotions.

These human inclinations all affect the progress of science, but the end result of the scientific process is one that is rational to the core. Science has an internal coherence and predictive power that is unique among human endeavors. Historian of science Gerald Holton identifies the "apparent contradiction between the seemingly illogical nature of actual personal discovery and the logical nature of well-developed scientific concepts."[17]

This stark contrast between the messy business of scientific discovery and the end result, which is the most accurate description possible of how the physical world operates given current knowledge, comes about in part because of a process known as "peer review." It is important for nonscientists to understand how this process works, because accurate peer review, more than any other part of the scientific process, determines whether a particular set of research results is credible. For a detailed look at peer review, go to <http://en.wikipedia.org/wiki/peer_review>.

THE IMPORTANCE OF PEER REVIEW

The formal process of peer review is most often conducted for articles submitted to scholarly ("peer-reviewed") journals or for research proposals submitted to funding agencies (such as the National Science Foundation). In journal reviews, the author submits the article to the editor, who then distributes the article to between one and four reviewers. The reviews are sometimes "blind" so that the reviewers don't know the names of the authors and their affiliations (it is however often possible for knowledgeable reviewers to infer the identities of the authors from a paper's citations and the authors' approach in presenting the results). More often, the authors' names are known to the reviewer. Only rarely is the *reviewer's* name revealed to the authors when comments are delivered. Both the number of reviewers and whether the reviews are blind are at the discretion of the journal editor.

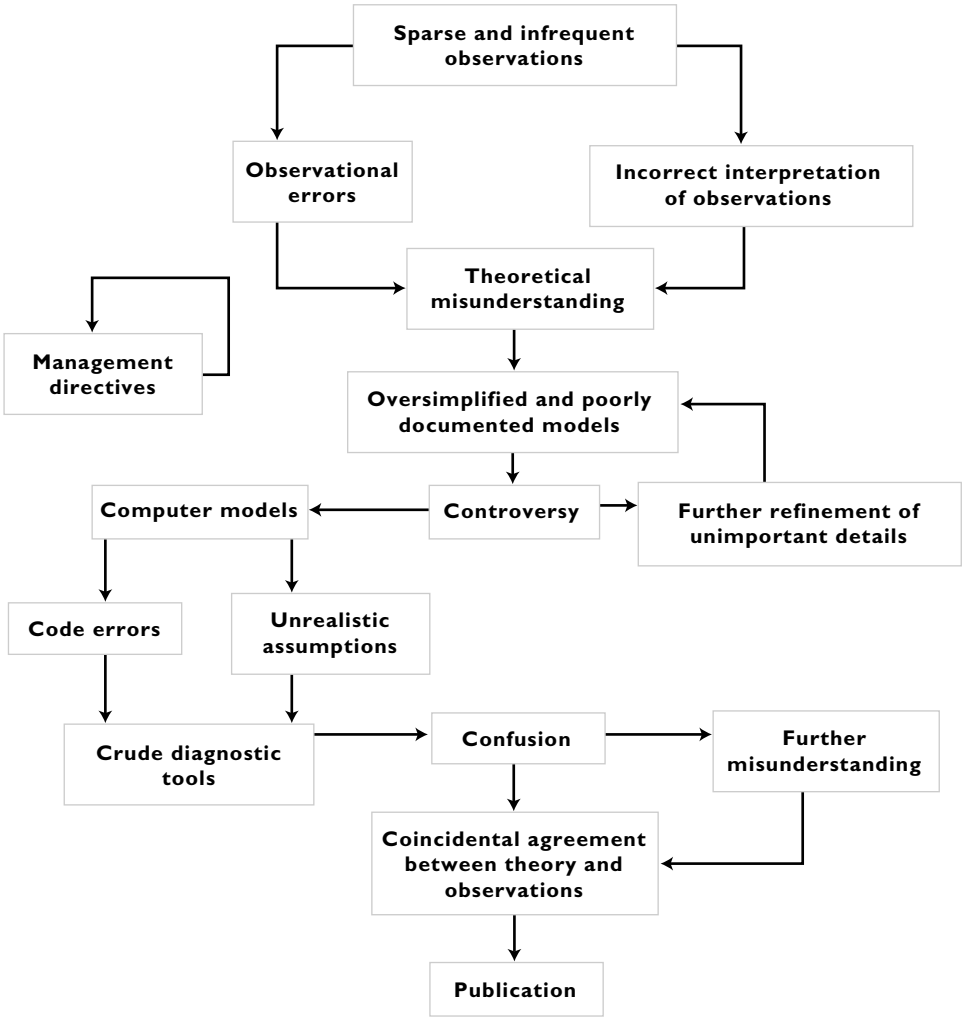
Reviews of scientific project proposals are usually more formalized than are journal reviews as befits a process upon which research funding depends directly. In addition, some journals and most funding agencies require researchers to declare any potential conflicts of interest before the review process begins.

The purpose of peer review is to judge whether the work is based on principles and judgments that represent the current scientific consensus. When such a consensus exists in a particular field, it reflects the paradigm accepted by scientists in that field and gives practitioners a standard set of tools with which to evaluate new ideas (see Kuhn's *Structure of Scientific Revolutions* for more details). When a paradigm is challenged by anomalous observations and experiments, a crisis ensues. The crisis continues until a new paradigm emerges that encompasses the anomalous results. During periods of crisis (or in non-scientific fields) peer review is generally less reliable because the foundations upon which the reviewers judge research results in the field are being questioned in fundamental ways.

Adequate documentation is a pillar of the peer review process. Without it, scientists could not verify or reproduce results, and scientific progress would grind to a halt. The best scientists are fanatical about documentation and you should be also.

Peer review does not guarantee accuracy. In fact, papers containing major scientific blunders have been published in peer-reviewed journals.[18] However, the formal peer review process makes it more likely that a paper will avoid major flaws. Be skeptical of "research" results that are not peer reviewed. There are some "scientists" who announce their results to the media but do not publish

FIGURE 4.1. A comical view of how human frailty can affect the course of science[19]



in peer-reviewed journals. Usually, such announcements are funded by particular organizations with an ax to grind and have little to do with science.

The "cold fusion" episode in 1989 is one example where the announcement of a "discovery" in a press conference but not in the refereed literature led to a media circus. Two scientists at the University of Utah (Martin Fleischmann and Stan Pons) claimed that they had created a room-temperature process that generated significantly more energy than it took to run the experiment, and that they had detected evidence that a nuclear fusion reaction was responsible. Had

the press been more skeptical of a research result announced only in a press conference, its pronouncements might have been more cautious. Instead, the claim of limitless cheap energy resounded loudly around the world, only to fall to the ground with a sickening "thud" when experiments by other scientists failed to corroborate the results.

How can you tell if a scientific study has been peer reviewed? **Table 4.1** lists some important peer-reviewed journals for key scientific fields. If the study is published in one of these journals, it's passed the first hurdle of basic scientific credibility. For a long (but still probably incomplete) list of peer-reviewed journals, check out <http://adswww.harvard.edu/abs_doc/refereed.html>. For a ranking of the importance of different journals in virtually all scientific fields, see the *Science Citation Index*, put out annually by the Institute for Scientific Information or ISI <http://www.isinet.com>. These rankings are based on the number of citations to papers in a given journal by the scientists in that field, so they are a relatively objective measure of journal quality (although no index is perfect and this one is no exception, as Wikipedia points out: <http://en.wikipedia.org/wiki/impact_factor>). In its electronic database, ISI's web site also has a large searchable list of journals, as well as other related products.

SCIENTIFIC REASONING

There are two kinds of reasoning accepted in scientific research. The first, based on deductive logic, relies on assumptions and general principles to derive implications and predictions of specific events.[20] If I leave my bicycle outside and it rains, I deduce from my experience and the laws of chemistry that parts of the bicycle will rust. Deductive inferences rely on a set of initial assumptions ("axioms"), which are analyzed using rules of logic. If the initial assumptions are correct, the rules of logic are followed, and no logical contradictions arise in the analysis, then the conclusions must be correct.

The second kind of reasoning–inductive logic, on which many inferences in science and other forms of human endeavors are based–relies on compilations of specific instances to infer general laws from the specifics (see Hughes, *Critical Thinking*). If I leave my bicycle unlocked outside for a hundred days and it is never stolen, I could use inductive reasoning to infer that it will not be stolen tomorrow, but I cannot be sure. I can only say that it is *unlikely* to be stolen,

TABLE 4.1. Some important peer-reviewed scientific journals

Field	*Journal*
Basic Science/Medicine	Science Nature Cell Nature Medicine
Clinical Medicine	New England Journal of Medicine Journal of the American Medical Association Annals of Internal Medicine
Physics	Physical Review Physical Review Letters Nature
Chemistry	Science Accounts of Chemical Research Journal of the American Chemical Society
Physical Chemistry	Journal of Physical Chemistry Journal of Chemical Physics Chemical Physics Letters
Biology	Science Nature
Geology/Geophysics	Journal of Geophysical Research — Solid Earth Geophysics Applied Geophysics
Global climate	Global Biogeochemical Cycles Tellus Biogeochemistry
Ecology	Ecology Ecological Applications
General Environmental Science	Environmental Science and Technology Atmospheric Environment
Hydrology	Water Resources Research Journal of Contaminant Hydrology
Combustion Technologies	Combustion and Flame Combustion Science and Technology
Electrochemical Technologies	Industrial Engineering & Chemical Research Journal of Power Sources Journal of Electrochemical Society
Physical Chemistry Technologies	Applied Surface Science Journal of Non-Crystalline Solids Applied Spectroscopy
Biological Technologies	Methods in Enzymology Biochemical Pharmacology American Journal of Physiology

based on my extrapolation of past experience. Inductive reasoning is less certain than deductive logic (in rare instances it can be *as* certain, but never more so).

The *Encyclopedia of Philosophy* defines inductive argument to include "all cases of nondemonstrative argument, in which the truth of the premises, while not entailing the truth of the conclusion, purports to be a good reason for belief in it."[21] In plain English, an inductive argument shows that the conclusion is highly probable but not demonstrably true in the same way that a correct deductive argument is. Induction is the link between mathematics, deductive logic, and our experience of the physical world.

Deduction by itself is impotent because it depends only on assumptions and its own internal logic. Without induction to enrich the set of data and assumptions upon which deductions can be based, the world of deduction would be a spare one, indeed. When inductive reasoning was first formalized, the goal of its devotees was to show that its results were demonstrably *true*, in the same way conclusions based on deductive reasoning were true. According to the *Encyclopedia of Philosophy,* "not until the end of the 19th Century did a more modest conception of inductive argument and scientific method, directed toward acquiring probability rather than certainty, begin to prevail." [22]

The limitations of induction are offset in part by the peer review process. Outside reviewers uncover errors of fact, logic, and omission, and report them to the authors and the journal. Obvious errors are weeded out quickly while more fundamental flaws may take years, decades, or even centuries to be uncovered. Eventually, though, even these errors will be discovered as scientists attempt to build on previous work. Serious flaws will result in logical inconsistencies (discovered through deduction) that will inevitably surface and be corrected.

THE SCIENTIFIC OUTLOOK

The physicist Alan Sokal points out that science is predicated on two key attitudes:

- being willing to accept what you find; and
- being willing to discover that you are wrong.

Human nature being what it is, these two attitudes are rare. Next time you find

yourself resisting a new idea, take a deep breath and try to see the other point of view. Ask yourself why this idea might make you uncomfortable (often it's
5 because it clashes with your ideology). If you can step back from a situation in this way, you will have achieved what I like to call the true "scientific outlook."

CONCLUSIONS

Science and technology are a critical part of modern life. You owe it to yourself to understand at least the basics of how scientific knowledge is created and used.

Remember always that science is a human endeavor. Learn about the process so you can better appreciate how to interpret scientific findings. Finally, always give more weight to peer-reviewed research than to results announced solely in the media.

EXERCISE

Find a newspaper story summarizing a scientific study that interests you. Find out where the author of the study works. Who funded the research? What's the name of the journal in which the research was originally published? Is it a refereed journal? Send for the report and compare it to the newspaper article's summary of its conclusions. Is the newspaper summary accurate?

The supreme task . . . is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction. There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them. — **ALBERT EINSTEIN**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

PART

II

BE PREPARED

Few people are so brilliant that they need never prepare, and even they would do far better with a little work beforehand. Preparation is one of the foundations for success in any field. The six short chapters in this section describe important preparation that any good analyst should consider. Prepare well, and you'll never regret it.

If I had eight hours to cut down a tree, I'd spend six hours sharpening my ax. — **ABRAHAM LINCOLN**

CHAPTER

5

EXPLORE YOUR IDEOLOGY

Anyone who states that ideology has no effect on decisions is lying, deluded, or woefully ill-informed. Everyone has an ideology whether explicit or not. Ideology provides a simplified model of the world that reflects our values, biases, and experiences. It helps people make decisions in the face of imperfect knowledge.

Anton Flew's *Dictionary of Philosophy* defines ideology as "any system of ideas and norms directing political and social action," indicating the key role it plays in mediating between ideas and choices. Ideologies are often explicitly political (libertarian, conservative, liberal) but may also relate to preferences about material things.

Even technical subjects can be affected by ideology. In late summer 1997, certain hackers with a libertarian ideology actually protected a recently released version of the Apple Macintosh operating system from being illegally copied (even booby trapping or removing copies from public newsgroups) because they wanted to strike a blow against Microsoft, one of the dominant companies in the computer industry.

Computer users each see themselves in ways that reflect their ideologies. Apple Macintosh users, for example, often see themselves as going against the conventional wisdom–resisting the crowd mentality by choosing a computer they regard as technically superior (in the late 1990s, Apple fed that perception in its advertising, associating, for example, Albert Einstein's unconventional brilliance with its products). Microsoft Windows users, on the other hand, stereotypically see themselves as realists who are adopting the standard because they perceive its complete dominance to be inevitable, because they believe it is cheaper, because they value the greater availability of software for Windows, or because they are forced to use it at work.

Ideology is one connection between the world of ideas and the world of choices. It impinges upon the Cycle of Action in the assessment of unknown or

unknowable aspects of events and in the collection and interpretation of data for events that can be measured. Ideology guides the assumptions and value choices you make in the course of any analysis, and it can be particularly helpful (or particularly misleading) in situations where your knowledge is limited.

19

There's no guarantee that your ideology, which has been developed in an iterative process between your predispositions and your experience, will work well in the future even if it's worked well in the past. That's why ideologies that maintain their relevance over time must allow for adaptation and evolution when confronted with new data and experience.

My own ideology, based on reason, pragmatism, environmental consciousness, and social responsibility, has not changed much during the past two decades. One core part of this ideology is, in the physicist Alan Sokal's words, "a respect for evidence and logic, and for the incessant confrontation of theories with the real world."[23] My ideology is constantly challenged by friends who find my rock-solid belief in reason distasteful, but thus far I have not been convinced to modify that part of my ideology in a significant way. Over the years, I have changed positions on particular issues, but my core belief in logic and evidence has never wavered.

The ideology held by an individual *can* change radically over the course of a lifetime. Eldridge Cleaver, one of the original Black Panthers in the 1960s, became a conservative Republican as he grew older. Such shifts can be triggered by a watershed event, or can happen gradually as a person's experience affects his or her outlook.

It is essential to think through your ideology so that it will be consistent. You can check its *external* consistency by examining how its assumptions and conclusions stack up against facts you know for sure. Sometimes there is no way to check external consistency, and we say in this case that the assumptions behind the ideology are not falsifiable, testable, or provable. We can always test *internal* consistency, however. If there are internal inconsistencies, we know there's something wrong.

To help think about your ideology, ask yourself these questions:[24]

- Where do you get your information?
- Whom do you trust? Why do you trust them?
- With whose opinions do you normally agree? Do you agree with all their

opinions? Why do you agree? Do you believe in their values or do you trust their judgment?
- To whom do you give the benefit of the doubt? Why?
- Are there events that outrage you? Why do they make you mad?
- What motivates you to work hard? From where did those motivations come?
- Could one part of your ideology lead you to take actions that would be inconsistent with another part of your ideology?

Ideology is the filter through which you view the world. By exploring your own body of beliefs, you can more reliably make choices that are consistent with your values, and you will also be better prepared to use numbers responsibly and effectively in your analyses, writing, and presentations.

EXERCISE

After you've asked yourself the questions above about your own ideology and distilled the answers into a coherent credo, try to answer the same questions for a few people whose ideology is quite different from your own (or ask them the questions directly!). This may help you understand them better, which is the first step toward true communication.

Only the supremely wise and the ignorant do not alter.

— **CONFUCIUS**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

6

GET ORGANIZED

Many people garner high scores on standardized tests but are unable to apply themselves in a way that gives results. *Raw intelligence* is the ability to think quickly and accurately; *useful intelligence* is raw intelligence as enhanced by hard work, common sense, instinct, experience, discipline, and organization. A person of moderate raw intelligence who possesses high useful intelligence will virtually always be more productive than a brilliant person who lacks the personality traits needed to make that brilliance useful in the world.

Organization is the single most important factor in creating useful intelligence, and it is a *learned* skill. By consciously becoming more organized, you can substantially increase your effectiveness in virtually any endeavor.

Julie Morgenstern, in her book *Organizing from the Inside Out*, gives sound and practical advice for getting organized. By following her systematic approach for assessing your goals and analyzing your current state of disorder, you can begin to get control of the chaos. After you've used Morgenstern's tricks and tips for assessing and addressing the situation, you can take the following additional steps to achieve effective personal organization:

- *Control information flows*: Decide which information streams are relevant and which are detrimental to your professional effectiveness. Cancel magazine and newspaper subscriptions except those you use regularly. Limit your use of the World Wide Web to get news updates–with few exceptions, checking more than once a day is a waste of time and effort. Set aside specific times to answer mail and email (there's usually no need to respond to emails as they arrive). Recycle junk mail without opening it.
- *Buy a paper or electronic organizer*: Track to-do tasks, appointments, addresses, phone numbers, and notes in an orderly way. Get rid of the myriad slips of paper that confuse and confound.

  A trick I recently discovered to keep track more effectively of to-do tasks

is to create a blank page with different types of actions (like "call," "do," "email," "fax," "investigate," "find," "visit") and to write down your tasks under the appropriate category (see **Figure 6.1**). This approach immediately creates a task-oriented hierarchy that helps you prioritize. I keep separate pages like this for work and home to-do tasks. If things are really busy, you can even keep separate pages for different projects.

- *Clean your office*: Throw out extraneous stuff and store rarely used books. Make frequently used materials easily accessible. Keep only current work on your desk and put other materials in drawers and filing cabinets. Identify those items that are important to you and make them accessible. Morgenstern says "keep what you use and what you love. Toss the rest." Spend fifteen minutes at the end of each day to tidy up your office, so you can start fresh in the morning.
- *Develop a filing system*: Filing is for retrieval, *not* storage. Create a filing system that helps you (or others) retrieve stored information in a timely fashion. If you can't find an article or electronic file, it's worse than useless because you waste time looking for it. Be selective about the items you file. **7**
- *Take time to reflect*: Make sure you give yourself at least a half-hour per day of solid thinking time, with no interruptions. Unplug the phone and turn off the radio. **21**
- *Prioritize tasks and focus on the essential*: Too often, our days are consumed by "urgent tasks that are not important." For at least part of the day, focus on the "important tasks that are not urgent."[25] **10**
- *Document, document, document*: Documentation creates a trail for you and others to follow after the fact. It is essential to any intellectual work, but is often neglected. By making it a priority, you ensure that your intellectual contribution will remain important and useful for years to come. **33**

You have control over each of these steps. Do them, and you'll make the most of every opportunity. Avoid them, and you'll never be as effective as you could be. Even the most creative people, who often view organization as inimical to their personal style, would benefit greatly from making it part of their daily routine.

EXERCISE

Think of three important goals in your life that you were unable to achieve because you were disorganized. Remember those missed opportunities when you feel overwhelmed by the clutter, and use those examples to motivate your cleanup.

EXERCISE

Copy the example to-do list (Figure 6.1) and give it a try in your daily life. Alternatively, create a similar list that more closely reflects your needs.

In human endeavor, chance favors the prepared mind.

— **LOUIS PASTEUR**

FIGURE 6.1. An Example Task List

Call	Date Due	Do	Date Due
Give/Deliver	**Date Due**	**Get**	**Date Due**
Email/Search on Web	**Date Due**	**Find**	**Date Due**
Fax	**Date Due**	**Investigate**	**Date Due**
Mail	**Date Due**	**Delegate**	**Date Due**
Talk to/Visit	**Date Due**	**Arrange**	**Date Due**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

7

ESTABLISH A FILING SYSTEM

A key part of being organized is creating a filing system that helps you retrieve stored information in a timely fashion. Such a system is critical for both printed documents and (increasingly) computer files. If you can't find a file, it doesn't matter if you have it or not, and looking for it distracts from other more vital tasks.

Be selective about the items you file. For years I clipped articles from newspapers on subjects in which I maintained an active interest, thinking that someday I might write papers on those topics. In most cases I never did, and all that effort was for naught. There was one example where the files came in handy for a research project, but I can only think of that one. I stopped clipping most newspaper articles years ago, but I still save a few articles and reports that I find particularly valuable for current projects.

The futurist Peter Schwartz also gave up an elaborate filing system because he never used it. His approach is one to emulate: "I concentrate on educating myself; on passing information through my mind so it affects my outlook; on tuning my attention as if it were an instrument." He recognizes that with this approach he may not be able to draw upon previously compiled materials, but that can be a benefit in disguise: "Sometimes I must go back and re-create all the research I did several years before. But that, in itself, is valuable, because in the fields I care about, the facts have changed since I last went to look for them." Moreover, he saves time by only filing those items most crucial to his work.[26]

Paper should pass through your hands only once on the way to the file, to the recycling bin, or to the outbox. You may never achieve this ideal in practice, but strive for it and you will save time and be more effective.

Never clip anything you can easily find later, especially if you can search for it electronically. As more newspaper articles and reports have become available on the Internet, my need for hard copy storage has decreased. Current events are well documented on the World Wide Web, and you need only call up one of the many competent search engines to get a nice listing of recent articles on par-

ticular topics. Many newspapers even offer CD-ROMS or web retrieval of older articles, which makes clipping articles even less worthwhile. Such electronic retrieval sure beats clipping and filing articles! If you need to create your own database of electronic clippings, there are many capable sheet scanners that will quickly and efficiently put these clippings on your computer.

All reports that my former group creates at Lawrence Berkeley National Laboratory are put on the World Wide Web, in a file format (Portable Document Format, or PDF) that will allow all types of computers to read it (Adobe distributes their Acrobat Reader Program for free to anyone who wants to download it, at <http://www.adobe.com/products/acrobat/readstep.html>). Others can then download these reports without the time and expense of having to send the hard copies. I also don't have to worry about where the copies of the printed reports are–I can just call them up in my web browser and view them. Often this is easier than hunting around in the filing cabinet for an old report. To supplement this electronic storage, I organized by subject area three-ring binders that contain our most recent reports, for ready reference.

Order files by active projects and by major subject areas. Also have file drawers for people, institutions, and for administrative issues. My "People" drawer holds files for particular individuals who work in my field (it is usually easier for me to remember who worked on a particular report than to remember the subject heading where I filed that report–having a file for individuals gets around that problem). The "Institutions" drawer is analogous to the People drawer. Create an archive for old projects that is in one of the less accessible filing cabinets, and put the most widely used "active" files near your workspace. Make your electronic file structure mirror that of your hard copy files.

Allocate one part of a file drawer to "Travel and events" and create one file folder per trip or event, with destinations and dates clearly marked on each tab. Keep financial, logistical, and substantive information for each event in these folders. Order them chronologically and save them for a long time. Years later, you may need to know details of a certain event for tax or other purposes.

A key trick to making any filing system work is to go through your files periodically to familiarize yourself with where things are. This occasional sorting will give you the opportunity to discard archaic materials and refile items that belong somewhere else. No filing system is ever finished. Periodic sorting helps keep your system current and useful.

The process of developing a filing system will teach you something about your work. To create such a system, you will need to analyze and categorize the kinds

of work that you do, and you will surely learn something from this effort. You will also need to think carefully about the kind of information you most often retrieve from your files. If others must also be able to access your filing system, you might adopt a slightly different organization than if you are the only user.

In a World Wide Web browser like Mozilla Firefox, Safari, or Microsoft Explorer, it is easy to record "bookmarks" for sites that you frequently visit. It's important to organize these well, so create a structure that parallels your hardcopy and electronic filing systems. These bookmarks are crucial to fast information retrieval.

Computerized reference databases like EndNote can make any filing system more effective. Each report is one entry in the database, and it is a trivial matter to add a sentence in the notes section of each reference entry that indicates the file drawer in which the hard copy is stored. These databases can be searched electronically, which can be a godsend when you're looking for a crucial report.

To preserve access to your stored articles and papers, DON'T lend them out to others unless this lending is recorded in a notebook or computer file devoted to this chore. I routinely reach for a document on my shelves only to find that someone has borrowed it. Such a written record can help you track down missing reports when this happens.

Most people borrow reports and then forget they have them. You should insist that someone borrowing a report copy the relevant pages (obeying copyright laws, of course) and return it to you promptly. This insistence will save you precious hours in the long run.

EXERCISE

List major topic areas for your current work. Compare these topics to the headings of your filing system. How well do they match up? Do the same comparison for your computer files, electronic folders, and bookmarks saved in your web browser. Is it time to reorganize things?

There's nothing remarkable about it. All one has to do is hit the right keys at the right time and the instrument plays itself.

— **JOHANN SEBASTIAN BACH**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

8

BUILD A TOOLBOX

A problem solver needs a toolbox as much as a machinist needs a lathe. My toolbox contains tricks and techniques for efficient and effective problem solving. Each person's toolbox is a product of education and experience—it grows and evolves over time.

No toolbox is ever complete because every problem has its own unexpected subtleties; the toolbox that worked for past problems may need to be updated to face new challenges. The learning process never ends.

LANGUAGE

Each field has its own language. You must at least develop working familiarity with basic concepts in a given field before attempting to decipher a problem. If you are delving into a field for the first time, obtain a textbook and begin your reading there. Textbooks conveniently summarize the core information in a given field and usually provide a nice glossary and lists of further references. Another good place to begin is the list of frequently asked questions (FAQs) on a particular topic that is available on the Internet for virtually any issue you can name <http://www.faqs.org>. Talk to people you trust who know something about the topic because they can often point you to the best reference sources and save you time.

23

CALCULATIONS

An important part of any analyst's toolbox is the ability to create simple back-of-the-envelope calculations at a moment's notice, using approximations and memorized data. Such calculations are a quick way to check the quantitative assertions of others and are invaluable in both business and science.

29

Darrell Huff's book, *The Complete How to Figure It*, is a beautiful compilation of hundreds of ways to calculate numbers important to daily life. Huff is the fellow who blessed the world with the book *How to Lie with Statistics*. Some of the major headings in the book, just to give you a flavor, are as follows:

Lifetime money strategy
Some personal things (like life expectancy)
Interest and saving
Investing
Other people's money
House planning
Building things (like additions to your house)
Your car
Travel
Conversions

The book is chock-full of specific examples (more than 350!) that you can copy when doing your own calculations. I recommend adding it to your toolbox.

Another nice source of math techniques is a book written by Clifford Swartz called *Used Math*. It is more advanced than Huff's book, and summarizes the basics of calculus and other essential mathematical tools, mainly for students of science. If you find yourself needing to refresh your knowledge of more advanced math, this book is a great place to start.

For an accessible and entertaining book about statistics, try Larry Gonick and Woollcott Smith's *The Cartoon Guide to Statistics*. Don't let its title fool you. It's a solid introduction to statistical concepts and methods.

CALCULATIONAL COMPUTER TOOLS

To investigate any particular software package, I recommend that you browse in your local bookstore. Writing computer books has become a cottage industry, and within weeks of the release of new software a book appears on the shelves to assist you with its use.

Spreadsheets and databases are the two most common computer tools used for calculations. The first spreadsheet was VisiCalc, which was created in the

late 1970s and was a great innovation compared to the programming languages that had been commonly used for business applications before that. I use spreadsheets every day. They are fine for simple analyses and powerful enough for moderately complex modeling tasks. They are also handy for graphing results and creating formatted tables for reports.

28

Databases are useful when you have large amounts of data that all have the same format. For example, if you have a mailing list of three hundred names, with addresses, phone numbers, and email addresses, a database is the right tool for organizing it. A database will allow you to export the data in any format. You may want names and addresses in one case, but names and email addresses in another. Once entered into the database, you can extract them any way you want. Spreadsheets can often serve as simple databases as well.

A more advanced set of tools is available for statistical analysis of data. These software packages can be quite expensive and are generally for expert use only. They are worth investigating if you need the horsepower and have the knowledge to use them effectively. They can help you divine relationships between factors that are related in a complex way. Many of these complex analysis functions are also available in spreadsheets, which can be useful for such analysis if the data sets aren't too large and the number of variables to analyze is small.

EXERCISE

Browse through the computer help section at the bookstore and read about software you don't currently use. Think about a problem that you're wrestling with now, and see if there's a way to apply some of these unfamiliar tools to solve it.

Applying computer technology is simply finding the right wrench to pound in the correct screw. — **ANONYMOUS**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

9

PUT FACTS AT YOUR FINGERTIPS

My college buddies used to object to my tactics in our numerous discussions about public affairs because I would be the first to pull out a reference book to check their sometimes wild assertions.[27] While your friends may also think it unfair, your discussions will be far more fruitful.

GENERAL

The single most important document for a U.S.-based researcher in virtually any quantitative field is the *Statistical Abstract of the U.S.* It is available every year from the U.S. Superintendent of Documents and other sources for around $40. You can also buy a CD-ROM with all the Statistical Abstract's tables in electronic form or you can download those data for free from the web. It's a treasure trove, containing about 1,500 tables summarizing everything from air pollutant emissions to population, average temperatures to voter registration, and hunting licenses to federal budget outlays. The CD-ROM version is especially nice because it allows easy access to tables in spreadsheet format, giving you a head start on your analysis.

The *Statistical Abstract* is the best source for items (such as population and economic activity) needed to normalize any data you get from other sources (for example, absolute growth in cement production over time is not easily understandable until you normalize it on a per capita or per dollar of GDP basis). It also contains the latest information on inflation over time, for use in converting any current dollar statistics to constant (inflation-adjusted) dollars. The data are taken from a variety of sources and are carefully documented. The footnotes to the tables will give you leads on sources that you never knew existed. The

27

33

Statistical Abstract is a great place to start any investigation, and you should have a recent electronic or hard copy version available at all times. All libraries will have the latest one. To order *The Statistical Abstract* or to download tables from it go to <http://www.census.gov/compendia/statab/>.

A companion to the *Statistical Abstract* is a good almanac. Although some data in these two sources overlap, the almanac contains many useful items that *The Statistical Abstract* doesn't have, such as zip codes, postal rates, tax rates, historical events, sociopolitical facts about states and countries, and other information that you didn't know you really needed (but you really do). There are many almanacs; the one I have is *The World Almanac and Book of Facts* <http://www.worldalmanac.com>. *Time Magazine, The New York Times,* and *The Wall Street Journal* also publish almanacs, some of which are more specialized than others. An almanac of some kind will be available in any bookstore or library, or you can access the Information Please Almanac on the web at <http://www.infoplease.com/>.

Many newspapers and magazines offer searchable archives of past issues on the web or on CD-ROM. Using such archives makes initial research on new topics relatively easy. *The New York Times*, for example, allows web readers to search the past 150+ years of archived articles and to see the results of the search in summary form <http://www.nytimes.com>. Many public libraries offer access to similar CD-ROMS and on-line databases, so check your local library before buying.

Another excellent source of data is a book by Mary Blocksma called *Reading the Numbers*. This book summarizes the meaning of numbers we encounter in everyday life, like the codes found on the side of tires or the conventions used in describing fluorescent light bulbs. It is unfortunately out of print but can be obtained at used bookstores, <http://www.alibris.com>, and some libraries.

A similar book that is more readily available is Andrea Sutcliffe's *Numbers: How Many, How Long, How Far, How Much*. The variety of data contained in this book is truly astounding. In one section there are estimates of safe food storage times in the refrigerator and freezer, and in another there is a list of the number of different types of instruments in a typical symphony orchestra. The major sections include money, food and nutrition, health, travel, daily life, transportation, time, weather, people and places, the universe, conversion

tables, and everyday math. The book contains more than 600 pages of data, and it's well worth having on your shelf.

Worldwatch, in *Vital Signs 2007–2008*, presents essential information related to resource consumption and environmental effects around the world. These data include pollutant emissions, freshwater resources, land use, population and health, food and agriculture, wildlife and habitat, energy, atmosphere and climate, and policies and institutions. The data in *Vital Signs* are also available electronically from the Worldwatch Institute <http://www.worldwatch.org>.

WEB/INTERNET DATA

More data appear on the Internet every day. It is an impossible task to report the latest developments in a book like this one (things change too quickly), but there are a few key sites of which you should be aware. All URLs and changes to them can be found at <http://www.numbersintoknowledge.com>.

- *The Library of Congress site* <www.loc.gov> is an invaluable source. You can search the library's databases with little effort. Many local libraries have also placed their catalogs on the Internet, which can save you a trip next time you want to do research.
- *The U.S. Bureau of the Census site* <http://www.census.gov> is beautifully designed and is a treasure trove of social, demographic, and economic data. You can search by key word, by place, by clicking on maps, and in many other ways. Most of the data on this site are free or available at nominal cost. This site also houses the on-line version of the Statistical Abstract of the U.S. <http://www.census.gov/compendia/statab/>.
- *The National Technical Information Service* (NTIS) is still, as of this writing, the official source of many U.S. government reports. NTIS sells almost three million titles, and you can search on its nicely designed web site <http://www.ntis.gov> to find the one you want.
- *The U.S. Department of Energy's Energy Information Administration site* <http://www.eia.doe.gov> is the right place to find energy and some environmental data for the U.S. The site also has limited information on certain international topics, such as nuclear power or energy use in other

countries. Most of the data on this site are free or available at nominal cost.

- *The U.S. Environmental Protection Agency's site* <http://www.epa.gov> contains huge quantities of environmental data, as well as access to the EPA's reports.
- *The Fed Stats site* is a treasure trove for U.S. economic and other data <http://www.fedstats.gov>. It allows easy access via the World Wide Web to data from more than 50 federal agencies.
- *The Economagic web site,* which has both U.S. and international economic data, is located at <http://www.economagic.com>. This site claims to offer access to 200,000 data files (many of them for free), with special services for subscribers (fees for subscribers range from small to moderate depending on the level of service).
- *The National Center for Health Statistics site* <http://www.cdc.gov/nchs> contains data and reports on U.S. health that can be downloaded for free.
- *The U.S. Geological Survey* produces a superb on-line atlas for the United States, which can be found at <http://www.nationalatlas.gov>. This site allows the user to map standard political boundaries, identify the locations of airports and dams, and overlay the routes of roads and railroads. It also contains maps of Super-fund sites, air pollution emissions, mining operations, industrial facilities, and selected data from the U.S. Census.
- *Frequently Asked Questions (FAQs)* exist on the Internet for almost any topic <http://www.faqs.org>. Because these FAQs are compiled by experts in each field, they are a great place to begin your inquiries. Use Internet search engines to find the FAQs that will help you most.
- There are hundreds of *newsgroups* on everything from Star Trek to statistical analysis. By posting an email to the relevant newsgroup, you can often get answers to questions from people all over the globe in a matter of hours or days. You can generally access these newsgroups using your web browser (consult your Internet service provider for details, or go to <http://groups.google.com>).

There's a real art to searching for information on the Internet. Each different search engine brings to you only a subset of the information actually available, so you'll need to use several to be certain you've done a thorough search. For

example, when I searched for the words "information overload" (in quotation marks) on the Google search engine, it returned about two million items. When I used Ask.com, it returned about 560 thousand. Not all of the items on the Google list were in the Ask.com list, and vice versa. Skill is also involved in narrowing down your search so that you have at most dozens of documents to sort through, instead of the thousands or millions that these engines commonly produce for you. Search engines try to rank their results by relevance, but the ranking algorithms also differ.

In recent years, Google has become the dominant search engine, followed by Yahoo and MSN. Also check out <http://dir.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web> or <http://www.searchenginewatch.com>, which give search strategies and details on specific search engines. Each search engine uses different algorithms and syntax for defining a search. It's worth mastering your favorite search engine's tricks so that you can find information quickly and efficiently.

CONTEXT

Sometimes in presenting the results of analysis, you'll want to know the intellectual history behind a certain concept. For example, if you are giving a talk about energy policy to a general audience, you might want to open with a brief summary of how the concept of energy was developed and how our modern conception differs from that common in the 1800s. In this circumstance, two essential works are *The Encyclopedia of Philosophy* and *The Oxford English Dictionary*. The former is helpful in tracing the genesis of a particular concept through the ages while the latter tracks the origins of particular words. Each of these sources can be useful in writing a speech or a paper and are both found in most libraries.

An alternative to the conventional *Encyclopedia of Philosophy* is *The Stanford Encyclopedia of Philosophy*, which is a web-based encyclopedia of comparable scope to the hard-copy version, on line at <http://plato.stanford.edu>. It is con-

stantly being updated and evaluated by its editorial board (they call it a dynamic encyclopedia). The creators of this new tool felt that the old way of creating an encyclopedia was too slow to capture the rapid developments in philosophical thinking (particularly developments related to scientific concepts). For a more general on-line encyclopedia, go to <http://www.wikipedia.org>.

A good book of quotations is essential to liven up your writing and presentations, and a standard dictionary and thesaurus will make your writing more precise. These are available in myriad incarnations in any bookstore. You can also search Roget's thesaurus at <http://humanities.uchicago.edu/forms_unrest/ROGET.html> or the Merriam Webster thesaurus at <http://www.m-w.com/thesaurus.htm>.

For many academic topics, excellent summaries of the current state of research are contained in *Annual Reviews*. These reviews are published for 27 different fields, including political science, public health, the environment, and the physical and biological sciences. They are often the best place to start when you're trying to learn about a new field.

CONVERSION FACTORS

For physical constants and conversions, there's nothing better than *The Handbook of Chemistry and Physics*. It comes out every year, but because physical laws and constants rarely change even an old version of this work is handy. Pick one up at a used bookstore. Virtually all libraries have a copy of this book on the shelf.

There are certain facts that I use frequently. For example, I often need to convert English units to metric units and vice versa. To help this process along, I memorize the most commonly used conversions and keep the rest on a one-page data sheet for ready reference. I recommend compiling such a sheet for your own use that contains the conversions most useful to your work. The hour or so of work needed to compile this page will be repaid many times over in the convenience of not having to dig deeply into obscure reference books whenever you need to do a calculation. See **Figure 9.1** for an example data sheet.

EXERCISE

Compile a data sheet for your own use after thinking for a week or two about the numbers you use most frequently. Pass it around to your co-workers to get their suggestions. Print copies for them when you're done, to share the wealth.

Data, Data, Data! I can't make bricks without clay!

— **SHERLOCK HOLMES**

FIGURE 9.1. Example Data Sheet

Physical Constants and Conversions (underlined values are exact)

Reprinted with permission from J. P. Holdren, January 2000.

1 ton (t) = 1,000 kg = 2204.6 lb = 1.102 short ton * 1 lb = 453.6 g * 1 oz = 28.35 g
1 atomic mass unit = 1.6606e-24 g * 1 grain = 6.480e-2 g * 1 carat (metric) = 0.20 g
1 m = 3.281 ft = 39.37 in = 1e6 micron = 1e10 angstrom = 1e15 fermi * 1 in = 2.54 cm
1 km = 0.6214 mi = 0.5400 naut mi * 1 ft = 0.3048 m * 1 parsec = 3.0857e16 m
1 m2 = 10.76 ft2 = 1e28 barn * 1 km2 = 100 hectare (ha) = 247.1 acre = 0.3861 mi2
1 mi2 = 2.590 km2 = 640 acres * 1 ft2 = 929.03 cm2 * 1 in2 = 6.452 cm2
1 m3 = 999.97 liter = 35.31 ft3 = 264.2 gal * 1 ac-ft = 1233 m3 * 1 km3 = 0.2399 mi3
(1 liter = 1 kg of H2O at 4 degrees C; density of H2O at this T is 0.99997 g/cm3.)
1 cord = 128 ft3 = 3.625 m3 * 1 bushel = 3.524e-2 m3 * 1 Imp. gal = 4.546e-3 m3
1 m/s = 3.281 ft/s = 2.237 mph = 1.944 knots * 1 mph = 0.869 knots = 1.467 ft/s
1 cfs = 0.0283 m3/s = 449 gpm = 724 ac-ft/yr * 1 m3/s = 35.3 cfs = 25.6e3 ac-ft/yr

1 J = 1e7 erg = 1 W-s = 6.241e18 eV = 0.7376 ft-lb * 1 Btu(IT) = 1055.1 J (Internat Table Btu&cal = 1.00067x thermo Btu&cal) * 1 cal(th) = 4.184 J, 1 cal(IT) = 4.187 J
1 kWh = 3.6 MJ = 3412 Btu(IT) = 859.8 kcal(IT) = 1.341 hp-hr * 1 hp = 550 ft-lb/sec
1 t TNT = 1e9 cal = 4.184e9 J (TNT explosive energy = 900-1100 cal/g)
1 a.m.u. converted to energy = 1.492e-10 J * 1 quad = 1e15 Btu = 1.0551e18 J
1 TW = 31.536 EJ/yr = 29.89 quads/yr = 14.12 Mbbl(US oil)/day = 1.076 Gt(UN coal)/yr

1 t coal equivalent (tce) [UN,WrldBnk,Shell] = 7,000 Mcal(IT) (low heat) = 29.31 GJ
1 t oil equiv (toe) [BP,Shell] = 10,000 Mcal(IT) (low ht) = 41.87 GJ —> 5.7 GJ/bbl
1 bbl crude oil [USDOE] = 5.8e6 Btu (IT, high heat) = 6.12 GJ —> 44.9 GJ/ton
1 t crude oil [UN] = 1.454 tce = 42.6 GJ * 1 t crude oil = 7.33 bbl = 1.166 m3
1 bbl petroleum products [USDOE] = 5.48 MBtu = 5.78 GJ * 1 bbl gasoline = 5.54 GJ
1 tce [BP] = 0.667 toe = 27.9 GJ = 26.5 MBtu * 1 t utility coal [DOE] = 24.76 GJ
1,000 m3 natural gas [BP,WorldBnk] = 9,000 Mcal(IT) (low ht) = 37.68 GJ; [Shell] = 9,500 Mcal(IT) (high heat) = 39.77 GJ; [UN] = 1.332 tce = 39.0 GJ * 1 t = 1,084 m3
1,000 ft3 nat'l gas [USDOE] = 1.028 MBtu -> 38.3 MJ/m3 * 1,000 kWhe [BP] = 0.25 toe = 10.47 GJ (low); [DOE] = 10.47 Mbtu (11.0 GJ) hydro, 11.02 Mbtu (11.6 GJ) nucl (high)
1 t U3O8 @ 0.5% utilization = 336e3 GJ —> 30.3 GWh(e) if converted at 32.5%
10 m3 water/sec x 10 m head = 0.98 MW * 1 t air-dried wood (1.39 m3) or dung = 15 GJ
1 t charcoal = 29.3 GJ (produced @ 25% effic. in dirt kiln, 50% in masonry kiln)
1 t crop wastes (20% moisture) = 13 GJ * 1,000 liter kerosene = 37.6 GJ
C02 emissions 25 kgC/GJ coal, 20 kgC/GJ oil, 15 kgC/GJ natural gas

1 newton (N) = 1e5 dynes = 0.2248 lb(force) * 1 pascal (Pa) = 1 N/m2 = 1e-5 bar
1 atm = 101,325 Pa = 10.33 t(force)/m2 = 14.696 lb/in2 = 760 mm Hg (0 C) = 760 torr
1 curie (Ci) = 3.7e10 dis/s = 3.7e10 becquerel (Bq) * 1 rad = 100 erg/g = 0.01 gray
1 gray (Gy) = 1 J/kg * 1 roentgen (R) —> 2.58e-4 coulomb(+ or -)/kg dry air
roentgen-equivalent-man (rem) = rads x quality factor * 1 sievert (Sv) = 100 rem

EARTH: R = 6357/6378(po/eq) km; A = 510.1e6 km2 (oc 361.3, ld 148.8, ice-free ld 133 Asia 50.0, Africa 30.3, N. America 22.4, S. America 17.8, Oceania 8.5, Europe 4.9);
M = 5.97e21 t (oc 1.4e18, ice 2.9e16, atm 5.2e15, lks/str 1.3e14, dry biom 1.8e12);
E @ top of atm = 172,000 TW (surf 88,000, evap 41,000, wind 2,000, GPP 200, NPP 100)
Highest elev = 8848 m; avg elevation = 840 m; deepest ocean = 11,000 m.
1 degree of latitude = 111 km; 1 minute = 1.85 km. Solar constant = 1368 W/m2.

ATM: @ sea lvl 1.225 kg/m3, 288 K, mol wt 28.96, 2.55e19 molecules/cm3, sound speed 331.4 m/s; percent by vol N2 78.09%, O2 20.95%, A 0.93%; ppm—CO2 340, Ne 18, He 5.2, CH4 2.9, Kr 1.1, N2O 0.5, CO 0.1, O3 0.01; atm water vapor & clouds 1.3e13 t.
Heat cap of air = 1 J/g-K. Tropopause: h = 11 km, p = 0.23 atm, T = 217 K. Wet-air adiabatic lapse rate = 6.5 K/km, dry = 10 K/km. Viscosity = 1.8e-4 dyne-sec/cm2.

H2O: heat fusn = 79.7 cal/g; heat cap = 1 cal/g-K; heat vapn = 539.6 cal/g (100 C), 584 cal/g (21 C), 597 cal/g (0 C). Ocn vol = 1.35e9 km3; avg depth = 3.7 km; mixed layer depth = 80 m; sea wtr density (15 C) = 1.026 g/cm3, salinity = 35,000 ppm (Na 10,500, Cl 19,000, S 885, C 28, N 15, Li 0.17, P 0.07, Fe 0.01, U 0.003, Au 0.00001)
Viscosity = 0.01 dyne-sec/cm2.

g = 9.807 m/s2 * R = 8.3144 J/mole-K * N = 6.022e23/mole * c = 2.998e10 cm/s
k = 1.3806e-23 J/K * s = 5.670e-8 W/m2-K4 * K = C + 273.15 * F = 1.8xC + 32
AMU = 1.661e-27 kg = 931.5 MeV * e = 1.602e-19 coulomb * h = 6.626e-34 J-sec

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

10

VALUE YOUR TIME

Mr. Casals, you are 95 and the greatest cellist that ever lived," a young reporter commented. "Why do you still practice six hours a day?" Casals replied, "Because I think I'm making progress."

Americans always seem to rush around. There's never enough time to do everything you want to do, see everyone you want to see, or travel everywhere you want to go. In part, this busyness is an inherent part of modern-day life, but it is also partly self-imposed.

Your time is your life, and it is precious. If someone is wasting it, they are stealing it from you, and YOU are allowing that theft. Many people don't think about this equation, and they allow people and institutions to monopolize their time and sap their energy without commensurate rewards. You must manage your time to reflect your priorities (no one else will do this for you).

The most articulate statement of this principle is found in *Your Money or Your Life*. In this book, Joe Dominguez and Vicki Robin describe a set of simple techniques to help people internalize this equation and make it part of their daily decision making. They encourage people to evaluate their spending by calculating how many hours of labor ("life energy") will need to be spent for each purchase. The value of money becomes more real after you ask yourself two key questions:

- "Did I receive fulfillment, satisfaction, and value [from this purchase] in proportion to life energy spent?"
- "Is the expenditure of life energy on this purchase in alignment with my values and stated life purpose?"

"No, Thursday's out. How about never—is never good for you?"

© The New Yorker Collection 1993 Robert Mankoff from cartoonbank.com. All Rights Reserved.

One hundred dollars has an abstract quality. It is made more real when comparing various purchases that can be made for that amount. It becomes most real when put in terms of the four or eight or twenty hours of labor needed to earn the money to buy this new toy you want.

This line of argument dates back to Thoreau, who, in *Walden,* compared the cost of train fare to the hours of time needed to earn the money to pay for it, and concluded that he would arrive in less time by walking. It has implications for your choices of where to work and how to spend your personal time.

Stephen Covey, in his book *The Seven Habits of Highly Effective People*, describes how most people allow urgent items to dominate their lives, but the most effective people focus significant attention on "*the important items that are not urgent.*"[28] Be selective about the endeavors on which you spend time. Prioritize. Choose jobs with the greatest likelihood for success and the highest potential payoff. Working at a dead-end job can sap time and energy from a more valuable use of your time.

Think carefully about time spent with others: Does it add to your life? Does

THANK YOU ALL FOR COMING. THERE'S NO SPECIFIC AGENDA FOR THIS MEETING...
AS USUAL, WE'LL JUST MAKE UNRELATED EMOTIONAL STATEMENTS ABOUT THINGS WHICH BOTHER US. I'LL KICK IT OFF...
THERE'S NEVER TIME TO GET ANY WORK DONE AROUND HERE!!
SAdams
© 1992 United Feature Syndicate, Inc.

DILBERT: © Scott Adams/Dist. by United Feature Syndicate, Inc.

it make you happy? If not, then do something about it. Don't let anyone steal your life!

Avoid going to committee meetings except where the agenda is well focused and the person chairing the meeting will keep things on track. Most meetings wander into irrelevance, and most people who call meetings have no respect for the time and money wasted. To avoid this problem, one executive prohibited meetings greater than ten minutes in length, with good results.

Avoid large meetings except when they are absolutely essential to improving the productivity of those attending. I knew my meetings were valuable when my staff of 12 actually requested that our biweekly group meetings last for a full hour instead of the 35 minutes we typically took, to allow more in-depth discussion of each person's current research. Of course, some meetings are held to foster teamwork or for political reasons, and these may not be so easily avoided. If you must attend, bring reading or other work so that, during the inevitable slow times in the meeting, you will be able to accomplish something.

You must protect your most productive time from interruptions. Each of us works best at particular times of day. I find that my best writing occurs in early afternoon and early evening. My best calculations are done in the late morning. A very small number of hours account for the bulk of my writing and research output, and if I don't write during those times, I'm vastly less productive.

Find your most productive times and guard them jealously. Unplug the phone. Go to the library. Take control of those times! Your life can be stolen by constant interruptions just as well as it can by people wasting large blocks of time in meetings.

In their classic book, *Peopleware*, Tom DeMarco and Timothy Lister write about how software companies can increase the productivity of their employ-

ees. Software designers are among the most creative people anywhere, and their productivity and accuracy are dependent in part on a workplace that isn't so noisy that it interrupts their work. *Peopleware* describes how programmers who found their workplace "acceptably quiet were *one-third more likely* to deliver zero defect work."[29]

Protect your work area from stray noise. If your office is close to others who are distracting you, perhaps you can negotiate with them. If that's not possible, then work in the library or at home. Don't force yourself to work under noisy conditions: you'll be wasting your time and your life.

If you are feeling unproductive, do something that takes little mental energy or effort but will yield dividends later (like sorting your files, reading mail, or answering phone calls). Even "down time" can be useful if you match the task to your mental energy level.

Finally, value other people's time just like you value your own. Don't schedule needless meetings. Rarely hold long meetings. Be prepared for all meetings. No one ever complains when meetings finish early.

EXERCISE

Keep a log tracking your time for two or three typical days, writing down the topics you worked on. Record all interruptions as well as the duration of uninterrupted time. Are you making the most of your productive hours? Are you surprised by how many interruptions there are?

Furious activity is no substitute for understanding.

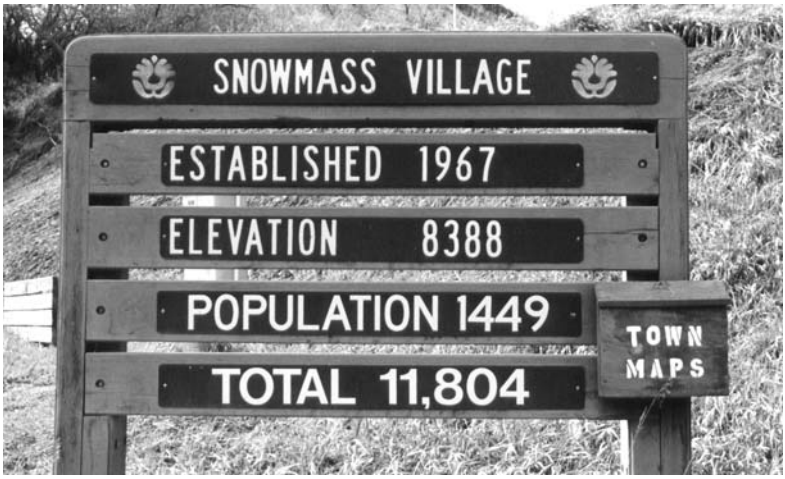
— **H. H. WILLIAMS**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

PART

III

ASSESS THEIR ANALYSIS



© 2001 by Jonathan Koomey. All Rights Reserved.

Any time you're exploring work others have done, you'll need to know what to look for. Your own systematic thinking can uncover logical blunders that the less critical observer will overlook. Hugh and Lillian Lieber (*The Education of T. C. MITS)* give many examples of apparently obvious conclusions that are erroneous, as does Darrell Huff in *How to Lie With Statistics*. William Hughes, in his book titled *Critical Thinking*, explores in a more general way correct and incorrect modes of argument.

By investigating carefully and thoughtfully, you will be able to evaluate other people's analysis as a matter of course and thus contribute to your understanding of the task at hand. The following chapters give some advice on how to make these explorations productive ones.

Just because I use a study to refute another study does not mean my study is right. It just means I believe it. Caveat Emptor.

— **CYNTHIA CROSSEN**

CHAPTER

11

THE POWER OF CRITICAL THINKING

Years ago, Dr. Norman Vincent Peale became famous for proclaiming the power of positive thinking. His claim was that each of us has control of how we feel about our lives, and, by using that control, we can create a better future for ourselves.

For assessing the analysis of others, *critical thinking* is at least as important as positive thinking. Critical thinking implies the ability to assess the true meaning of arguments and the skill to determine whether those arguments are credible and compelling. William Hughes, in his excellent book *Critical Thinking*, defines some of the key terms for understanding this important skill:

> When we express an inference in words, we do so by means of statements. A *statement* is a sentence (i.e. a set of words) that is used to make a claim that is capable of being true or false.... An *argument* is a set of statements that claims that one or more of those statements, called the *premises*, support another of them, called the *conclusion*. Thus, every argument claims that its premises support its conclusion.[30]

A *logically strong argument*, according to Hughes, is one "whose premises, if true, support its conclusion." A *sound argument* is "a logically strong argument whose premises are true." Critical thinking is the way to identify and create sound arguments, which is the main goal of all analysts.

Hughes describes the key steps a critical thinker uses to assess arguments and decide whether they are sound.[31] I summarize his steps below:

- "*Identify the main conclusion*": Rephrase the main conclusion in your own words, to be sure you understand it. Think about the context in which the argument is being made. Finally, if the argument is not stated clearly, interpret the argument in the way that is most charitable to your opponent's views (Hughes calls this "the principle of charity").

- "*Identify the premises*": Rephrase the premises in your own words. List the unstated premises and assumptions. Discard distracting and irrelevant examples that are not really premises. Don't forget that context is important, and that the principle of charity should guide your interpretations at this stage as well.
- *Identify the logical structure of the argument and determine how the premises support the conclusion*: Determine whether one premise depends on another one, or, alternatively, if premises are independent of each other. Determine whether the argument contains sub-arguments that support the ultimate conclusion. Hughes presents a way of describing the structure of arguments, known as a tree diagram, that can help with this step.
- *Ask whether you'd be justified in accepting the premises*: In most cases, a strict proof of the premises is impossible, so we must be satisfied with the less stringent standard of "acceptability." Hughes suggests two criteria for evaluating acceptability:
  - — If the statement is common knowledge, we should regard it as acceptable, unless the context requires a higher standard of proof.
  - — If the statement is not common knowledge, we should ask for, or be prepared to offer, the evidence upon which it is based, and accept it only if the evidence meets the appropriate standard, for example, personal experience, appeal to a recognized authority, or strict proof.

14 16 Hughes also points out that "common knowledge" is often wrong, so use care when relying on any measure of acceptability less than a true proof.

Key questions to ask when determining acceptability include: Are the premises common knowledge? What is the context of the argument? Who 14 is making the argument? Are they experts? Do they have a vested interest in 4 the outcome? Are the premises backed up by credible research published in 25 peer-reviewed journals? Are the premises based on consistent comparisons?

- *Ask whether the premises are relevant*: Many arguments are peppered with premises that are irrelevant to the conclusion but appear to support it. Usually the previous steps will make it obvious which premises don't apply, but it's worth asking the question again at this stage just to be sure.
- *Ask whether the premises "adequately" support the conclusions*: Certain statements lend some support to the conclusion but are not sufficiently 14 compelling to make the argument adequate (appeals to authority often fall into this trap). Inadequacy is a key weak point of many arguments. This

step is mainly an exercise in common sense and in figuring out whether a particular argument accounts for all of the many possibilities.

- *Look for missing arguments:* Omitting a valid comparison is one commonly used way to make an argument seem stronger, so try to find these omissions in other people's arguments. People often look for the best-case comparisons to support their own position and ignore comparisons that might weaken their case.
- Finally, "*Look for counter arguments*": Try to find a sound argument that leads to a conclusion contradictory to the one reached by the argument being evaluated. If you find such a counter argument, you know that something's wrong, because two sound arguments cannot contradict each other (unless one is not truly sound).

Careful critical thinking is at the root of all good analysis. When these steps become second nature, you will have mastered its essence. The many subtleties of assessing arguments will always keep you on your toes though. There's no end to the ways that arguments can be flawed.

EXERCISE

In a newspaper, find a letter to the editor on a topic you care about. Identify the conclusion and the main premises. Assess each premise, determining whether each is acceptable, relevant, and adequate to support the conclusion. How easy was it to carry out Hughes's process?

The fascinating impressiveness of rigorous mathematical analysis, with its atmosphere of precision and elegance, should not blind us to the defects of the premises that condition the whole process. There is perhaps no beguilement more insidious and dangerous than an elaborate and elegant mathematical process built upon unfortified premises.

— **T. C. CHAMBERLAIN**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

12

NUMBERS AREN'T EVERYTHING

Don't confuse things that are merely countable with those that really count.[32] Some analysts systematically ignore things that can't be turned into a number and assume that because these things can't be measured, they are unimportant. These analysts also generally assume the converse, that things that can be measured must be important. Both mistakes are common, especially among the quantitatively inclined, but these are errors to which you need not succumb.

Four categories of things affect decisions in the world:

- Some things can be measured directly and with confidence before decisions have to be made. The amount of money in your bank account at the time of your last statement falls into this category.
- Some (perhaps most) things are not directly measurable but can be estimated accurately enough in time to influence decisions. You cannot know exactly how much money you will have in your bank account in a month, but you can estimate it in a crude way by making assumptions about your expenditures, based on an extrapolation of past spending trends and planned expenses.
- Some things cannot be measured, estimated, or predicted accurately in time to influence decisions. Some reasonable estimate might be possible for such factors, but time is often limited. For example, time constraints may prevent you from doing a complete analysis of your assets before rushing to put a bid down on a house.
- Some things cannot in principle be measured, estimated, or predicted under any circumstances. Sometimes there are physical reasons for this inability (e.g., the Heisenberg uncertainty principle limits the accuracy of measurements of the velocity and position of subatomic particles) and at other times we cannot measure or predict because things that really matter to people's lives (such as loyalty, love, honesty, and integrity) are sim-

20

What's the value of the earth?

YOU'RE PACKING?
YEP. GET YOUR TOOTHBRUSH, HOBBES. WE'RE OUTTA HERE.
IT'S AN OUTRAGE. HOW GROWN-UPS HAVE POLLUTED THE EARTH! I REFUSE TO INHERIT A SPOILED PLANET! I'M LEAVING!
REALLY? WHERE TO??
YOU KNOW, SOMETIMES YOU'RE A REAL LOAD TO HAVE AROUND.
I WAS JUST ASKING.!
© 1988 Universal Press Syndicate
WATTERSON
9-13

CALVIN AND HOBBES © 1988 Watterson. Dist. by UNIVERSAL PRESS SYNDICATE. Reprinted with permission. All rights reserved.

ply not quantifiable in any meaningful way. It is not possible to estimate the true value of a warm summer's day or a species saved from extinction except in terms that have nothing at all to do with numbers.

Events in the last two categories are typically treated in the Evaluation stage of the Cycle of Action (I call these unknown or unknowable aspects of events). To analyze such situations, we must make assumptions and value judgments, using our ideology and instinct as guides to fill in the facts that we can't obtain by measurement or estimation. **3** **19** **5**

Items in the last category are commonly ignored but often are of the most importance to human life. Because many things simply can't be quantified, we can only insist that decision factors that defy quantification be listed and described in as complete a form as possible, given current knowledge. Then, at least, decision makers are aware of the other factors that may relate to the decision at hand and can weigh these factors based on values, instincts, experience, ideology, and knowledge. If decision makers persist in ignoring the intangibles, then the situation has moved beyond the realm of analysis, and an explicitly political debate about the value of those intangibles must then ensue. **5**

EXERCISE

Think again about the three recent decisions that you explored in the Cycle of Action chapter. Which of the data affecting those decisions could be measured directly? Which could be estimated and which could not? How did you treat the intangibles that could not be quantified? What assumptions did you make to treat the unknowables?

To laugh often and much; to win the respect of intelligent people and the affection of children; to earn the appreciation of honest critics and endure the betrayal of false friends; to appreciate beauty; to find the best in others; to leave the world a bit better, whether by a healthy child, a garden patch or a redeemed social condition; to know even one life has breathed easier because you have lived. This is to have succeeded. — **BESSIE A. STANLEY**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

13

ALL NUMBERS ARE NOT CREATED EQUAL

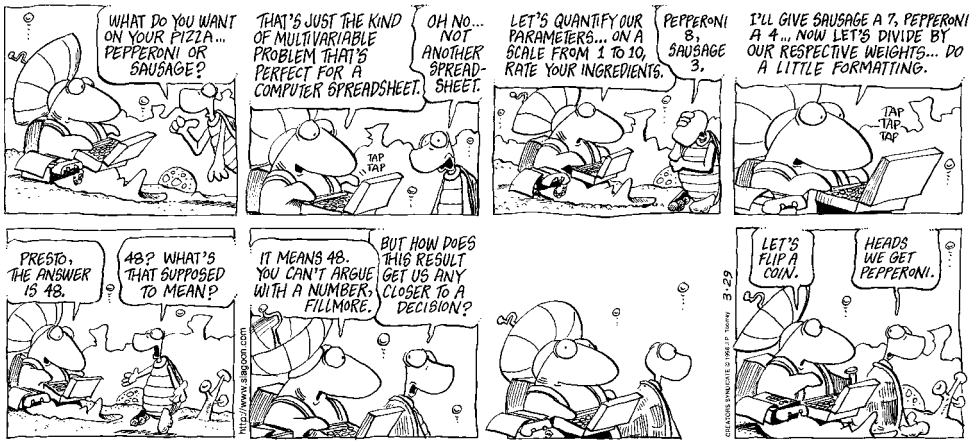
A common fallacy to which even sophisticated analysts sometimes fall prey is to look for immutable laws of human behavior, in much the same way as the physicist discovers physical laws through experiment. Such generalizations about human and economic systems fail because these systems are adaptable in ways that physical systems are not.

Physicists conduct replicable experiments to uncover fixed natural laws. A scientist measuring the speed of light in a vacuum, for example, would find the velocity the same in the U.S. or Tahiti. If another scientist conducted a similarly accurate experiment in Russia, that test would produce the same result.

On the other hand, relationships between cause and effect for individuals and human institutions are dependent on institutional, social, and economic context. Furthermore, these relationships change over time. A market researcher attempting to predict consumer acceptance for a new toothpaste would find that a market test in San Francisco would likely yield quite different results than in Tahiti even though the speed of light remains the same in both places. In addition, if the same experiment had been conducted in the 1950s, the results would have presumably varied wildly from those of the current day.

In the mid-1970s and early 1980s, conventional wisdom held that modern societies could not reduce energy use without also reducing gross domestic product (GDP) and harming their economies. For example, Gordon Corey, vice-chairman of Chicago's Commonwealth Edison, stated in 1981 that "there is an unbreakable tie between economic prosperity and energy use." Similarly, the Chase Manhattan Bank stated, in its 1976 *Energy Report,* that

> there is no documented evidence that indicates the long-lasting, consistent relationship between energy use and GDP will change in the future. There


WHAT DO YOU WANT ON YOUR PIZZA... PEPPERONI OR SAUSAGE?
THAT'S JUST THE KIND OF MULTIVARIABLE PROBLEM THAT'S PERFECT FOR A COMPUTER SPREADSHEET.
OH NO... NOT ANOTHER SPREAD-SHEET.
TAP TAP
LET'S QUANTIFY OUR PARAMETERS... ON A SCALE FROM 1 TO 10, RATE YOUR INGREDIENTS.
PEPPERONI 8, SAUSAGE 3.
I'LL GIVE SAUSAGE A 7, PEPPERONI A 4... NOW LET'S DIVIDE BY OUR RESPECTIVE WEIGHTS... DO A LITTLE FORMATTING.
TAP TAP TAP
PRESTO. THE ANSWER IS 48.
48? WHAT'S THAT SUPPOSED TO MEAN?
IT MEANS 48. YOU CAN'T ARGUE WITH A NUMBER, FILLMORE.
BUT HOW DOES THIS RESULT GET US ANY CLOSER TO A DECISION?
LET'S FLIP A COIN.
HEADS WE GET PEPPERONI.
3-29
http://www.sherman.com
CREATORS SYNDICATE © 1992 J.P. Toomey

SHERMAN'S LAGOON © JIM TOOMEY, KING FEATURES SYNDICATE

is no sound, proven basis for believing a billion dollars of GDP can be generated with less energy in the future.[33]

Believers in an unbreakable link between energy use and GDP assigned the immutability of a physical law to this historical relationship but found their belief shattered by events. From 1973 to 1986, U.S. primary energy consumption stayed flat, but GDP rose 35% in real (inflation-adjusted) terms. These believers had forgotten that people and institutions can adapt to new realities, and historically derived relationships (like the apparent link between energy use and GDP that held up for more than two decades in the post-World-War-II period) can become invalid when events (like the 1973 oil embargo) overtake them.

Most economic computer models embody historical experience through relationships that are derived statistically. Such models are often used to assess the potential effects of proposed changes in government policy or business strategy, but they cannot give an accurate picture of a world in which the fundamental relationships upon which they depend are in flux. If the statistically-derived relationships embedded in such a model are the very ones that would be affected by choices or events, then those relationships must be modified in the analysis; otherwise, the results are suspect.

For example, some economists trying to assess the costs of reducing carbon emissions assume that the only way to reduce these emissions (there are many) is to impose a large carbon tax that would raise the price of coal, oil, and natural gas in a way that has no historical precedent. They further compound this

error by using models that embody historical relationships to do long-term forecasts without altering those relationships to reflect the unprecedented nature and size of these taxes or the likely effects of other policy choices (see DeCanio 2003).

Creating a world with vastly lower carbon emissions presupposes massive behavioral and institutional changes that render past relationships between energy use and economic activity largely irrelevant (just like what happened after 1973). It also presupposes alterations in government policies. It is laughable to use computer models based on historical relationships for one-hundred-year forecasts in the face of such changes. Instead, scenario analysis is a more **26** appropriate tool for exploring the key relationships and how the future might unfold if those relationships change.

If you are confronted by results of computer-generated forecasts, always ask **30** whether the analysts changed the key historical relationships to reflect the possibility that those relationships will be affected by institutional choices or other developments. If not, then you've identified a key failing of the analysis and the results are suspect.

I have found that many physical scientists, computer modelers, and economists are susceptible to this mistake. I am not certain why forecasters from these disciplines fall into this trap, but I have noticed it often. It may be what my social-science-oriented friends call "physics envy," or it may be that most analyses are conducted in a mechanical way without significant reflection. In any case, once forewarned, you need not let them get away with it.

It is fashionable today to assume that any figures about the future are better than none. To produce figures about the unknown, the current method is to make a guess about something or other—called an "assumption"—and to derive an estimate from it by subtle calculation. The estimate is then presented as the result of scientific reasoning, something far superior to mere guesswork. This is a pernicious practice that can only lead to the most colossal planning errors, because it offers a bogus answer where, in fact, an entrepreneurial judgment is required. — **E.F. SCHUMACHER**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

14

QUESTION AUTHORITY

Sacred cows make the best hamburger. — **MARK TWAIN**

Question Authority" is one of the favorite T-shirt and bumper sticker slogans in Berkeley, California. It dates back to the 1960s when students engaged in well-publicized defiance of authority figures in the campus administration and the state government. Although some aspects of these emotion-charged times are best left in the past, it is time to dust off "Question Authority" for the 21st Century.

Of course there are excellent reasons for following authority in many instances. People can become authority figures by virtue of hard work, unique experience, special insight, and long training. In some situations, such as military maneuvers, obedience to authority is essential to completing missions successfully. Following authority in a business setting can also be important and can allow people to act in a concerted and collective way to achieve the common goal of increased profit for a firm.

However, most situations in everyday life are not as dangerous as military maneuvers. There are two main reasons why questioning authority is often a healthy response:

- Even brilliant people affiliated with well-respected institutions are not always right. History is littered with examples of wrong-headed conventional wisdom and misjudgments by otherwise outstanding individuals. Recall what a Yale University management professor said about Fred Smith's paper proposing reliable overnight delivery service: "The concept is interesting and well formed, but in order to earn better than a "C," the idea must be feasible." Smith went on to found Federal Express Corporation,

which in 2007 did more than 35 billion dollars in business carrying overnight packages.

- Many salespeople and marketing professionals use deference to authority to gain advantage and influence consumer behavior. Robert Cialdini, in his book *Influence*, documents dozens of amusing and sometimes disturbing examples of how these "compliance professionals" use authority figures and pseudo-authority figures to sell products. Cialdini describes one famous example where the actor Robert Young, who used to play Marcus Welby, M.D. on television, advises people to buy Sanka coffee for its health benefits: Cialdini concludes "objectively, it doesn't make sense to be swayed by the comments of a man we know to be just an actor who used to play a doctor. But, as a practical matter, that man moved the Sanka."

So long as there is time to reflect and consider the dictates of authority, it's healthy to question it. Hughes, in his book *Critical Thinking*, proposes five criteria to use when assessing the assertions of an ostensible authority figure.[34] I summarize them below:

21

11

- "*The authority must be identified*": Appeals to anonymous authorities are immediately suspect. You can't assess the strength of an appeal to authority without knowing who the authority is.
- "*The authority must be generally recognized by the experts in the field*": At a minimum, the authority must possess an academic degree, a professional license, or an institutional position that indicates acceptance by the relevant expert community.
- "*The particular matter in support of which an authority is cited must lie within his or her field of expertise*": The authority must be a real expert on the topic at hand, but it is common for experts in one field to express opinions about another field, and be taken seriously whether their opinions are well founded or not (the Robert Young example described above is a case in point).
- "*The field must be one in which there is genuine knowledge*": Hughes defines genuine knowledge as "a systematically ordered body of facts and principles that are objective in nature." Beware of appeals to authority in fields that don't fall under this definition.
- "*There should be a consensus among the experts in the field regarding the particular matter in support of which the authority is cited*": Even in

physics, there are issues that remain a subject of controversy. Appeals to authority are only relevant on topics that are not controversial.

11 Unless these criteria are satisfied, an appeal to authority is not *adequate* to support a particular argument and should probably be disregarded.

Cialdini also suggests asking one more question, namely "how truthful can we expect the expert to be here?" By asking whether the authority figure (or the institution funding her) stands to benefit from the advice she is supplying, we gain insight into the trustworthiness of that advice. If experts have a vested interest in one or more potential outcomes of your decision, their advice should be discounted and you should seek corroborating information. The advice may nonetheless be sound, but it is worth special care in this case.

If the expert's interests coincide with yours or the expert is demonstrably impartial, then more weight should be ascribed to his or her opinions. Even in this case, however, it is best to seek corroborating evidence because experts can be mistaken.

One recent example where the interests of the researchers and their funding institutions led to widespread (and well-justified) public skepticism is the furor over research on the health effects of smoking. For years, the tobacco companies funded research ostensibly demonstrating that smoking did not cause adverse health effects, all the while knowing that it did (for details, see *The Cigarette Century*). Investigating who is paying for the expert's work is a particularly fruitful line of inquiry when such cynical manipulation of research is so common.

Some authority figures unknowingly promote the messages of self-interested institutions. In *Diet for a New America*, John Robbins marvels at the way that the dairy and meat industries have for decades promoted attitudes favorable to their products by contributing educational materials to schools. All of us who grew up in the 1960s and 1970s remember the advice from our teachers to drink milk and eat meat to promote good health, but those nice charts describing the basic "food groups" were part of a brilliant marketing campaign, in spite of how official they looked.

The flood of facts and falsehoods in this data-laden age puts greater responsibility on us to question "information" that appears to be from authority fig-

ures. Information found on the Internet is less subject to institutional validity checks than are stories published by more established information sources like *The New York Times* or *The New Republic*, but any source can propagate nonsense. Proceed with caution, and always corroborate what you hear or read with multiple independent sources before taking action.

16

Loyalty to petrified opinion never yet broke a chain or freed a human soul. — **MARK TWAIN**

CONVENTIONAL WISDOM (SIC) FROM WELL KNOWN AUTHORITIES

"But what . . . is it good for?" —Engineer at the Advanced Computing Systems Division of IBM, 1968, commenting on the microchip

"There is no reason anyone would want a computer in their home." —Ken Olson, president, chairman and founder of Digital Equipment Corp., 1977

"So we went to Atari and said, 'Hey, we've got this amazing thing, even built with some of your parts, and what do you think about funding us? Or we'll give it to you. We just want to do it. Pay our salary, we'll come work for you.' And they said, 'No.' So then we went to Hewlett-Packard, and they said, 'Hey, we don't need you. You haven't got through college yet.' " —Apple Computer Inc. founder Steve Jobs on attempts to get Atari and H-P interested in his and Steve Wozniak's personal computer

"This 'telephone' has too many shortcomings to be seriously considered as a means of communication. The device is inherently of no value to us." —Western Union internal memo, 1876

"The wireless music box has no imaginable commercial value. Who would pay for a message sent to nobody in particular?" —David Sarnoff's associates in response to his urgings for investment in the radio in the 1920s

"Who the hell wants to hear actors talk?" —H.M. Warner, Warner Brothers, 1927

"We don't like their sound, and guitar music is on the way out." —Decca Recording Co. rejecting the Beatles, 1962

"Heavier-than-air flying machines are impossible." —Lord Kelvin, president, Royal Society, 1895

"Professor Goddard does not know the relation between action and reaction and the need to have something better than a vacuum against which to react. He seems to lack the basic knowledge ladled out daily in high schools." —1921 *New York Times* editorial about Robert Goddard's revolutionary rocket work

"Stocks have reached what looks like a permanently high plateau." —Irving Fisher, Professor of Economics, Yale University, 1929

"Airplanes are interesting toys but of no military value." —Marechal Ferdinand Foch, Professor of Strategy, Ecole Superieure de Guerre, early 1900s

"Everything that can be invented has been invented." —Charles H. Duell, Commissioner, US Office of Patents, 1899

"Louis Pasteur's theory of germs is ridiculous fiction." —Pierre Pachet, Professor of Physiology at Toulouse, 1872

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

15

HOW GUESSES BECOME FACTS

All data should be treated with skepticism. Here is one example of how official statistics get created, as recounted by Alan Meier, a long-time friend and colleague of mine at Lawrence Berkeley National Laboratory (LBNL):

> In 1987, Steve Greenberg and I wrote an article in an energy magazine about the rising amount of energy use that did not fit into the traditional categories.[35] As part of the article, we created three tables showing ownership of these small appliances (like fish tanks and power tools) and their estimated annual energy consumption. These values were based on very limited monitored data, back-of-the-envelope calculations, and hunches. The tables were assembled in one evening. (Many of the envelopes with calculations were then discarded.)
>
> In 1989, the U.S. Department of Energy's Energy Information Administration (EIA) published its official *Household Energy Consumption and Expenditures*, including the results of their 1987 survey and additional analysis.[36] The EIA published the Meier and Greenberg data in a new table, "U.S. End-Use Consumption of Electricity for Selected Appliances." Whereas we published ranges in our estimates, the EIA just calculated and printed the averages from the high and low values. The word "estimated" appeared nowhere in the EIA table, so the reader was led to believe that these numbers were exact (curiously, the EIA is careful to give confidence bounds and other statistical parameters for its own survey data). To add insult to injury, the EIA misspelled my name in the citation.

29

Alan's example is more the rule than the exception. When little information is available about a particular topic, any moderately credible source gets cited by everyone concerned with the topic and becomes the new conventional wisdom. This happens frequently even though such estimates are often based on extremely crude assumptions.[37] I now sometimes joke that ALL estimates of energy use of appliances somehow originate with Alan.

Another colleague, Chris Calwell, was hired in 1996 to write a research report about the energy and safety problems with halogen torchieres, those inexpensive floor lamps that have become so popular in recent years.[38] Here is his story:

> In order to write the report, I needed to figure out how many halogen torchieres had been sold. The trouble was, nobody knew. The Census Bureau knew how many halogen bulbs had been imported but not how many were in fixtures. California utilities had counted how many were in a small sample of houses, but that was before the lamp had become popular. Market researchers had asked how many people bought halogen lamps of all types but didn't know how many were torchieres. The library was not much help, either, because only two articles had ever been published on this subject before I set out to write mine.
>
> I called the author of those two articles and got the names and numbers of her sources (the manufacturers). Then I called the manufacturers and asked how many they thought had been sold in total the previous year. I also asked them how average prices for the lamps had changed over time and about when they began selling the lamps in the United States. Taking all these different pieces of information under consideration, I created a table of sales of torchieres per year and the number of fixtures still in use at the present time. It was pure guesswork, informed by the information I could find, but guesswork nonetheless, with a fancy spreadsheet and graphs to back it up.
>
> The U.S. Consumer Product Safety Commission, a federal agency, was about to put out a press release on the problems with halogen torchieres at the same time I was finishing my report, but they also had no idea how many had been sold. They asked for my number (40 million), which my publisher reported to them as 35 to 40 million to be conservative. The CPSC thought that range seemed high, so it used 30 to 40 million in its press release to be even more conservative. During the next year, dozens of newspaper articles and TV programs cited the CPSC estimate of 30 to 40 million and attributed it to the federal agency, ignoring the original source and never bothering to examine the eight-line footnote in my original report documenting how the original estimate of 40 million was determined. The number had taken on a life of its own, proof of truth by institutional adoption and media repetition.

33 Your best defense against these sanctified guesses is to read the documentation for any data, including the footnotes. Track down cited sources and read them,

too. You should be able to figure out the methods used to create any data. If the documentation is not up to this task, you should regard the data with extreme suspicion. Don't ever use data unless you know how they were derived, you trust the cited sources, and you agree with the stated methods.

27

EXERCISE

Find an official statistic that sounds plausible and explore its origins. Do you still find it plausible after you've investigated?

The Government are very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases.

— **SIR JOSIAH STAMP,**
Inland Revenue Department of England,
1896-1919

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

16

DON'T BELIEVE EVERYTHING YOU READ

Never act solely on information you read in the newspaper, hear on the radio, or learn on the Internet. News stories are often wrong either because the reporter misunderstood her sources or because the sources themselves were misleading or incorrect. Reporters sometimes report fiction as fact. Email claiming to contain valuable information can easily be a hoax.

One example is that of Stephen Glass, who wrote about 40 articles for *The New Republic* in the mid-1990s. Almost three-quarters of those stories were "partly or totally bogus," and the editor of that magazine was forced in mid-1998 to print a contrite mea culpa when Glass's deceptions were discovered.[39] The standard fact-checking process was not equipped to detect deliberate deception.

Some reporters even attribute quotations to those they have interviewed, when the person being interviewed never said any such thing. Erik Brynjolfsson, a professor at MIT's Sloan School of Management, had this experience in the spring of 1999. After I showed him a newspaper article in which he was cited, he said "I think I actually said about 25% of the quotes that were attributed to me. At least they didn't make me say anything too embarrassing." No one knows how frequently this happens, but it's common enough that you should not assume the veracity of quotations until you've verified them with the actual source.

Sometimes the sources themselves are at fault. Through the mid-1990s, the Software Publishers Association (SPA, now known as the Software & Information Industry Association, <http://www.siia.net>) periodically released estimates of the size of the software markets for Microsoft Windows and Apple Macintosh systems. In 1995 and 1996, the SPA proclaimed that Macintosh software sales had declined in both years relative to the prior year's estimates. The problem is that the SPA compared *preliminary* data in 1995 and 1996 to *final*

data in the previous year, but the final data in a given year contained substantial adjustments relative to the preliminary data (in the 1995 data, the adjustment was 45%!). The data series used by SPA was therefore not a consistent one, but the SPA's analyst felt confident enough to conclude that "the Mac platform is in decay." In fact, if final 1995 numbers are compared to final 1994 numbers, the real change in Mac software sales was about a 25% increase from 1994 to 1995, not a 14% decline as initially reported by the SPA. This result is not surprising, because Apple shipped record numbers of Macs in 1995.

The details of why the SPA made these appalling errors are not really that important for our purposes. The main lesson is that even respected institutions can distribute junk data to the world and pass it off as legitimate. If some Mac zealots had not delved into these numbers and compared the results year to year, unsuspecting software developers might have been led to the wrong conclusion. In fact, these widely reported estimates probably damaged Apple's business. How much damage they did no one can say for sure.

In the face of the public furor over its obvious mistakes, the SPA finally suspended its data program in September 1997. In hindsight, the SPA should have subjected its methodology to external peer review and been much more cautious in its conclusions about Macintosh software sales.

4

The most famous book on ways to misuse numbers is Darrell Huff's *How to Lie with Statistics*. Huff documents dozens of examples where people, through incompetence or malicious intent, mislead readers with numbers, graphs, and other tools of the numerate. These tricks include biased samples, carefully chosen averages, graphs with chopped scales, and faulty logic in all its many manifestations.

You owe it to yourself to read Huff's book. It is justly revered both because of its accessible tone and its message: lying with statistics is commonplace, and it's wise to be aware of the most widely used tricks.[40]

What should you do to protect yourself and others from such shenanigans?

- First, *go back to the questions*. Read the questionnaire. Find out who asked the survey questions and how they settled on the survey sample. Also determine when the questions were asked, because that can sometimes affect people's responses.

17

- *Be a detective* and determine the potential motives of the surveying organization. Identify the funding sources for the research.

24

27

- *Dig into the numbers*, and compare results to other numbers you know to be true. Are results consistent with other sources? Can you find internal inconsistencies that cast doubt on the results? Answers to these latter two questions made it obvious that the SPA results were nonsense.
- Finally, don't email information to dozens of people unless you are confident it is true. I've received dozens of email warnings about computer viruses that will erase my hard drive or have some other terrible effect. I know most of these are bogus because they usually don't mention which kinds of computers they affect, and all genuine warnings are very clear about the software that is affected by a given virus. Viruses (with a few exceptions) afflict Macs, PCs, or UNIX machines, but one type of machine can't generally "catch" the virus from another type. Before forwarding ANY email about computer viruses, check out the latest information at the Symantec Anti-Virus Research Center, <http://www.symantec.com/avcenter/index.html>.

Because of the ease with which information can be sent to dozens of people, news (or falsehoods) can spread quickly around the world. One example of this phenomenon is a collection of advice purported to be a commencement speech given by Kurt Vonnegut at MIT. The advice was witty, thoughtful, and altogether in keeping with Vonnegut's curmudgeonly but lovable take on the world. In fact, his own wife passed it around to her friends, thinking it was his work. It was widely distributed on email in the days following the MIT Commencement on 5 June 1997, and attributed to Vonnegut.

The essay included such gems as:

> Wear sunscreen. If I could offer you only one tip for the future, sunscreen would be it. The long-term benefits of sunscreen have been proved by scientists, whereas the rest of my advice has no basis more reliable than my own meandering experience. I will dispense this advice now.
>
> Do one thing every day that scares you.
>
> Sing.
>
> Don't be reckless with other people's hearts. Don't put up with people who are reckless with yours.
>
> Be kind to your knees. You'll miss them when they're gone.
>
> Live in New York City once, but leave before it makes you hard. Live in Northern California once, but leave before it makes you soft.
>
> Floss.

The only problem was that Kurt Vonnegut didn't speak at that MIT Commencement, nor did he write those delightful words of wisdom. The essay was a newspaper column by Mary Schmich, who writes for the Chicago Tribune. The column had somehow been stamped with Vonnegut's name and then sent merrily on its way to thousands of unsuspecting email readers.

This last example is a reminder of how quickly and easily misinformation can be passed off as fact in the digital age. Remember it as you venture forth into cyberspace.

EXERCISE

Find three popular articles on a topic you know a lot about. Make a list of the various misstatements you can identify based on your knowledge of the topic. Now go and read a few other articles on topics you know less about. Assume you find just as many mistakes. Worrisome? You bet!

The surest way to corrupt a youth is to instruct him to hold in higher regard those who think alike than those who think differently.

— **NIETZSCHE**

WHAT SCIENTISTS SAY, AND WHAT THEY MEAN: A FACETIOUS GUIDE FOR THE PERPLEXED

PHRASE	TRANSLATION
It has long been known	I haven't bothered to look up the reference
It is believed	I think
It is generally believed	A couple of other guys think so, too
It is not unreasonable to assume	If you believe this, you'll believe anything
Of great theoretical importance	I find it kind of interesting
Of great practical importance	I can get some mileage out of it
Typical results are shown	The best results are shown
3 samples were chosen for further study	The others didn't make sense, so we ignored them
The 4-hour sample was not studied	I dropped it on the floor
The 4-hour determination may not be significant	I dropped it on the floor, but scooped most of it up
The significance of these results is unclear	Look at the pretty artifact
It has not been possible to provide definitive answers	The experiment was negative, but at least I can publish the data somewhere
Correct within an order of magnitude	Wrong
It might be argued that	I have such a good answer for that objection that I shall now raise it
Much additional work will be required	This paper is not very good, but neither are all the others in this miserable field
These investigations proved highly rewarding	My grant is going to be renewed!
I thank X for assistance with the experiments and Y for useful discussions on the interpretation of the data	X did the experiment and Y explained it to me

SOURCE: Internet lore.

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

17

GO BACK TO THE QUESTIONS

It is common practice for companies to make decisions about marketing new products based on surveys of consumer behavior and preferences. In 1996, I spent about four hours filling out such a survey, and it got me to think about the limitations of such data.

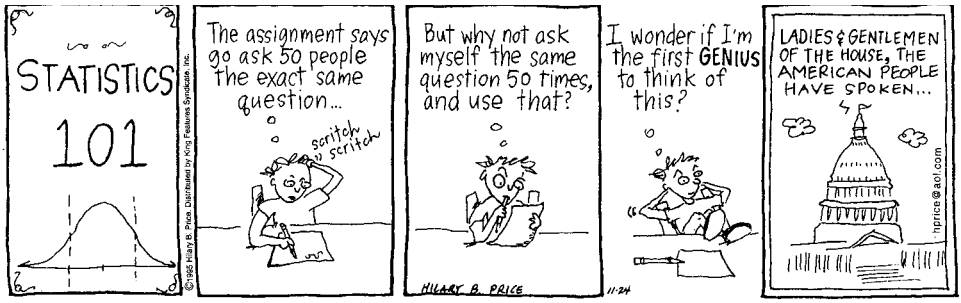
The Survey of American Consumers is conducted by the American Institute of Consumer Studies.[41] An engaging young woman came to my door, explained what her job was, and started asking me questions. After about 15 minutes of questions she said that she had a booklet with more detailed questions, and that if I filled it out and returned it in the postage-paid envelope, I would be sent a check for $20.

The survey itself was 96 pages long, and it contained questions on every aspect of my purchasing behavior, from what kind of toothpaste I buy to how many miles I drive my car per year. The sheer endurance needed to fill out the survey was substantial, and I suspect that I was among the more thorough respondents. I dug through my cabinets to identify the brands of various foods and other household supplies. Sometimes I just didn't have the brand name but I knew I had used some form of the product during the specified time period, so I guessed (the instructions to the survey said to "estimate as best you can").

One thing that struck me was the number of times that the question did not allow me to explain my situation. For example, the survey asked about my newspaper reading habits ("Which sections of the daily newspaper do you read?") but did not allow me to explain that I read the news on-line and did not at that time receive a hard copy newspaper.

Another question asked whether I use talcum powder. I do use it, but not to freshen up after a shower. Instead, I put it on ants that have invaded my kitchen; it drives them away, without using poison.[42] The makers of talcum powder might try to infer something from the results of this survey, but it tells them nothing about my non-conventional use of that product.

A worst case scenario...


STATISTICS
101
The assignment says go ask 50 people the exact same question...
scritch scritch
But why not ask myself the same question 50 times, and use that?
I wonder if I'm the first GENIUS to think of this?
LADIES & GENTLEMEN OF THE HOUSE, THE AMERICAN PEOPLE HAVE SPOKEN...
©1998 Hilary B. Price. Distributed by King Features Syndicate, Inc.
HILARY B. PRICE
hprice@aol.com

RHYMES WITH ORANGE © HILARY B. PRICE, KING FEATURES SYNDICATE

Still another question asked where I do my food shopping but did not give me an option for the two local food stores at which I shop. Thus, it appears from my responses to the survey that I had only gone food shopping once in the previous 30 days even though I had eaten quite well during that period, thank you very much.

Given a sufficiently large number of respondents, a survey like this one will probably give results that reflect typical consumer purchases. Nevertheless, the anomalies I encountered should at least make you cautious about drawing firm conclusions from one survey without independent confirmation of the results.

These cautions are especially important when using "survey" data that can be skewed by so-called "self-selection bias." This problem commonly crops up when the institution conducting the survey has no control over who responds, which is typical for surveys conducted over the World Wide Web. Respondents with a particular point of view can easily "stuff the voting box," as *Byte Magazine* found when it tried to assess the prevalence of different computer operating systems.[43] The *Byte* survey results so clearly did not reflect the reality of computer ownership in the U.S. that the magazine discarded the results.

In spite of all these limitations, millions of business investment dollars are spent based on the results of such surveys. Before making important decisions based on survey data, you should examine the actual questionnaire used to collect the data and not simply accept the data as they are reported. The way a question is asked can affect results in substantial ways. This problem is accentuated if the data are broken down to show their locations within census tracts or zip codes because the small number of respondents inside each of those boundaries may lead to unrepresentative results.

The ways in which questions are asked, and to whom, can determine the out-

come of a poll. Crossen, in her book *Tainted Truth*, recounts a classic example from the early 1990s:

> Consider this question posed by Ross Perot. In a mail-in questionnaire published in *TV Guide*, the question was "Should the President have the Line Item Veto to eliminate waste?" 97% said yes. The same question was later asked of a sample that was scientifically [randomly] selected rather than self-selected, and 71% said yes. The question was rewritten in a more neutral way—"Should the President have the Line Item Veto or not?"–and asked of a scientifically selected sample. This time only 57% said yes.[44]

To truly understand the meaning of a survey, it is essential to go back to the original survey questions. *Never* base critical decisions on someone else's summary of survey results unless you have implicit trust in that person's judgment and understanding of the situation. It is also crucial to understand how the sample of respondents was selected (self-selected samples are the bane of a good analyst's existence). The sample must be truly *representative* of the population or else the results are suspect.

Real data always reflect the inherent inability of humans to track the changes happening all around us. That's why it's important to check new data with intuition, experience, and independent sources of confirmation before taking action. **3**

EXERCISE

If you encounter results of a poll that you find completely unbelievable, contact the firm that conducted the survey and ask to obtain the survey questionnaire. Try to determine whether misleading or unclear questions are the cause of the bizarre results.

Everybody believes in public opinion polls, from the man in the street to President Thomas E. Dewey. — **GOODMAN ACE, 1948**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

18

READING TABLES AND GRAPHS

When reading a new report, I go straight for the tables and graphs. I do this for two reasons:

- I want to determine whether I find the author credible, and
- I like to understand the bottom-line results.

These two goals are quickly served by rummaging through tables and graphs for half an hour. If the tables and graphs are good enough, I read the paper to follow the author's reasoning more closely.

If the tables and graphs are poorly designed or confusing, I lose respect for the author. It is essential that tables and graphs summarizing analysis results be
33 clear, accurate, and well documented. If they aren't, including them is worse than useless because they hurt your argument and your credibility.

First check for internal consistency. I always begin at the bottom-line table and work backwards. I examine the column and row headings to be sure I understand what each one represents (I read the footnotes if I have questions).
27 I then assess whether the components of the total add up to the total. This procedure shows me whether the calculations are accurate, and it helps me become familiar with the different parts of the analysis.

Next, look over the numbers in the table and identify those that are abnor-
27 mally small or large. Typographical errors are quite common in tables (particularly in tables that summarize results from other, more detailed tables), and a quick scan can help you find them. For example, if you're reading a table summarizing hours worked per week by different team members on a project, an entry from one person that's ten times larger than the entries for others should catch your attention and prompt you to investigate further. The number may not be wrong, but checking it will increase your confidence in the analysis and help you understand it more fully.

Sometimes numbers don't add up exactly because of *rounding errors*, not because there is a mistake in the calculations. Say you've formatted a spreadsheet table so that there are no decimal places for the entries in the table. These entries might be 9.4 and 90.4, but in the table they are shown as 9 and 90, because a common convention is to round numbers down to the next whole number when the decimal remainder is less than 0.5 (the remainder in both cases is 0.4 in this example). The sum is 99.8, which rounds to 100 and is shown in the spreadsheet as the total. The sum of 9 and 90 is 99, which makes the total of 100 look wrong even though there's a perfectly sensible explanation. If you're not aware of this potential pitfall, you could be misled by these apparent errors.

Read the footnotes carefully. They should convey the logic of the calculations and the sources of key input data. If you can't determine the methods used from the footnotes, you should be especially suspicious of the results and investigate further.

Check for ambiguous definitions and terminology. For example, there are at least five distinct definitions of the word "ton," and analysts often neglect to specify which definition they are using. If it isn't crystal clear what the label means, you are likely to be led astray in interpreting the numbers. Another slightly different example involves the number of hours in a year. Many analysts assume it is 8,760 hours, but on average it is 8,766 hours because of leap years. This difference is essentially a definitional one, but it can lead to small inaccuracies in calculations. 33

The next step is to check consistency with independent external sources to ensure that the data in the tables and graphs are roughly right (use back-of-the-envelope calculations as needed). Compare growth over time to growth in other key data like population and Gross Domestic Product to get perspective on how fast something is growing relative to these commonly used indicators. 29 27

Does the information in the tables or graphs contradict other information you know to be true? That table of hours worked may list Joe as someone who worked little on the project at hand, but if you know for a fact that Joe slaved over this project for many weeks because it was his idea, then you'll need to check the calculation. Similarly, if there's an entry of 175 hours for one week, it must be a typo or a miscalculation, because there are only 168 hours in a week.

Take ratios of results and determine whether the relationships they embody make sense. If one component of the total is growing much faster than other 27

components over time, or if it is especially large compared to others, then investigate further. Look for large discrepancies and investigate when you uncover them.

Follow up when you encounter cognitive dissonance–any contradiction between your knowledge and the information in the table will lead to greater understanding, one way or the other. If there is a logical explanation for the contradiction, you've learned about the relationships between the information in the table and what you knew before. If the contradiction indicates a real problem, you've identified a flaw in the analysis. Root out the causes of cognitive dissonance, and you will enhance your knowledge without fail.

EXERCISE

Find several reports that you've either used or written over the past few years, and dissect the tables and graphs. Can you find all the information you need to use and to reference the results in the reports? How many unexplained discrepancies did you find? Have your tables and graphs improved over the years?

Although we often hear that data speak for themselves, their voices can be soft and sly.

— **FREDERICK MOSTELLER, STEPHEN E. FIENBERG, and ROBERT E. K. ROURKE**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

19

DISTINGUISH FACTS FROM VALUES

Every human choice embodies certain values. When technical people advocate a technology choice (e.g., whether or not to build more nuclear power plants) they often portray their advice as totally rational, completely objective, and value free. This portrayal *cannot* be correct—if an analyst makes a choice, he has also made a value judgment. For example, Bernard L. Cohen[45] portrays his support for a revival of nuclear power as the result solely of scientific analysis when in fact it is a *public policy judgment* that is based both on his values and his scientific understanding.

Robert Reich's public policy students often forgot that values (and hence politics) are always intertwined with public policy choices:

> Some of my students used to regard public policy-making as a matter of finding the "right" answer to a public problem. Politics was a set of obstacles which had to be circumvented so the "right" answer could be implemented. Policy was clean–it could be done on a computer. Politics was dirty–unpredictable, passionate, sometimes mean spirited or corrupt. Policy was good; politics, a necessary evil.[46]

One purpose of analysis is to support public or private choices; in this context, it cannot be "clean" or value-free. It can, however, illuminate the *consequences* of choices so that the people and institutions making them can evaluate the alternative outcomes.

When Victor Fuchs et al. analyzed economists' positions on policy issues, they found that those positions were strongly correlated with each economist's values. Economists with a conservative political ideology viewed policy issues one way while those with a more liberal bent generally viewed them in a different way. Jonathan Marshall summarized the results nicely:

> What the authors found was that economists' views on policy questions correspond more closely to their values than to the [facts].
>
> To some extent, in other words, mere facts don't change their mind on issues. Keep that in mind the next time you see an ad, signed by 200 economists, supporting some candidate or piece of legislation. The ad may tell you less about their expertise than their ideology.[47]

I define facts as assertions about the physical world that can be verified through experiment, direct measurement, or observation. Events can also be facts although the interpretation of the meaning of those events is usually strongly colored by ideology and attitudes. Values, on the other hand, are explicitly subjective and are an expression of the ideas and feelings that are most important to us. As the philosopher Ken Wilber points out, the only way to truly understand values (or other beliefs) in all their complexity is to engage in a dialogue with the person who holds them.

5

Understanding the relationship between facts and values can aid decision-making.[48] If people agree about both the facts and the values related to a particular decision, reaching an acceptable solution becomes merely an issue of computation. A situation in which people concur about values but disagree about facts presents another easy choice: in this case, experimentation is the order of the day. If the facts are not in question, but people's values are in conflict, then they can negotiate to solve the problem. The most complicated and difficult choices are those where both facts and values are in dispute. In such situations, paralysis or chaos results. Most difficult public policy choices are problems of this type, and the best approach here is to experiment and bring the facts to a point where people no longer dispute them; people can then start to negotiate about values.

In any of these discussions, it is a common rhetorical trick to link an ostensibly demonstrable fact with a value judgment. Such statements are of the form "Fact A is true, therefore we should take Action B." By linking facts and values in this manner, the speaker is explicitly or implicitly hiding the value judgment she has made.

Let the reader beware: Fact A may or may not be true, and Action B may or may not be a good idea whether Fact A is true or not. Too often these glib statements are allowed to pass without proper scrutiny. By being aware of this com-

monplace rhetorical technique (and by using your critical thinking skills), you can prepare yourself to respond. **11**

The purpose of analysis is to help you illuminate the facts so that you can understand the choices of others and best use your values to make good choices of your own. Value choices are inescapable, which is why separating demonstrable facts from value judgments is so essential. Do so, and you are sure to be on a solid footing.

The significant problems we face cannot be solved by the same level of thinking that created them. — **ALBERT EINSTEIN**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

20

THE UNCERTAINTY PRINCIPLE AND THE MASS MEDIA

One of Werner Heisenberg's major contributions to quantum theory was the so-called "uncertainty principle." This principle governs the relationship between the observer and the physical world at the subatomic level. To measure the velocity and position of a particle (an electron, for example) an observer must bounce something off that particle but by that action will change the electron's velocity and location. The more the observer tries to pinpoint the location of the electron, the more the measurement disturbs the velocity. Similarly, if the observer attempts to measure the velocity of the electron, its position becomes less and less certain. The level of certainty with which the observer can know the position and velocity of the electron is constrained by this physical law.

This principle can be loosely summarized for use in the everyday world as "the observer disturbs the system." Any time a measurement is made, that measurement changes the system being measured, usually in some miniscule way. Measuring the temperature of a liquid, for example, requires that a small amount of heat be extracted from the liquid and transferred to the thermometer. In physical terms, a good measurement is one that disturbs the system as little as possible. Things are a bit more complicated with social systems, especially when observations by influential members of the mass media can disrupt such systems in a significant way. The recent historical record is rife with both intentional and unintended disturbances of this kind.

The phenomenon of publicly available ranking systems is one example of how making new information available can lead to institutional change. Before 1984, when Milton Moskowitz and Robert Levering first wrote their book *The 100 Best Companies to Work for in America*, people had difficulty learning about family issues related to job choices because systematically compiled information was not yet available. When this work was published, the rankings became a force for change that had not previously existed.[49] The same effect

occurs whenever pollution emissions or campaign contributions are made public in an accessible way: embarrassment, education, or enlightened self-interest cause people and institutions to modify their behaviors (check out the Scorecard web site at <http://www.scorecard.org> for a good example of how data and analysis tools can be designed to affect institutional behavior). 38

Another example of intentional disruption of observed systems is the set of opinion polls that influence the opinions they attempt to measure.[50] The pollsters ask questions of a carefully selected sample of individuals and then create a press release summarizing some of the key results. These "promotional surveys" are increasingly common. They can be conducted by a company for its own marketing, a non-profit institution for its own fundraising, or by a "neutral" third party like a trade magazine. For example, Kiwi Brands (a shoe polish company) polls people on a regular basis on such important topics as the relationship between ambitious people and shiny shoes and what the biggest grooming mistakes are: not surprisingly, badly shined shoes come out near the top.[51] Another company well known for this tactic makes sugarless gum ("Four out of five dentists surveyed prefer . . ."). In any case, read the fine print and learn who was surveyed, how and when they were surveyed, and what questions they were asked. It is easy to create a survey that will give the desired result and there's no way to know the truth without going back to the survey questions. 17

When the mass media report on companies wrestling with financial problems, they often cause unintended harm. Reporters repeat again and again the time-worn opinions of pundits, both well- and ill-informed. A whirlwind of bad publicity creates uncertainty and doubt among a company's customers, thereby influencing the very event on which the news people are reporting. The reporters create, in Cialdini's words, a kind of "social proof" that a company will fail, thereby making it more likely that it will.

An unintended result of polls can be that their results, taken out of context, often make headlines. People read the headlines and rethink their opinions based on them (if "thinking" is the right word to describe this process). The poll itself becomes news, and, in so doing, it influences the outcome of elections and changes public opinion.[52] The exact effect polls have on elections is uncertain, but clearly they can influence the results, as shown by the 1948 U.S. Presidential election. Crossen quoted Burns Roper (son of the founder of the Roper polling organization) as saying that the 1948 polls "lulled the Republican supporters

into a false sense of confidence and scared the bejesus out of organized labor." The pollsters also believed that people had decided on their presidential choice weeks before the election, so they stopped regular polling. In addition, telephone polls only reached people with telephones, who were generally wealthier than the population at large and were therefore more likely to vote Republican. One famous result was the erroneous *Chicago Tribune* headline "Dewey defeats Truman!"

Mass media amplify the effect of such observations, and this tendency will only become more prevalent as the cost of transmitting information continues to plummet. Corroborate mass media reports with independent information before using them or passing them to others. If you repeat such reports without analysis, you enlighten no one and may cause harm to people and institutions that you do not intend. Finally, do your utmost to ensure that your own research results are reported correctly so that you yourself do not fall prey to this pitfall.

Public opinion is a compound of folly, weakness, prejudice, wrong feeling, right feeling, obstinacy, and newspaper paragraphs.

— **ROBERT PEEL**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

PART

IV

CREATE YOUR ANALYSIS

Now that you've gleaned what you can from the analysis of others, it's time to roll up your sleeves and get to work. Each of these short chapters offers key insights to help you get started. The order in which you read them is not so important, so first choose the ones that interest you most, but do read them all before you're through.

It is a mark of educated people, and a proof of their culture, that in addressing any problem they seek only so much precision as its nature demands or its solution requires.

— **ARISTOTLE, in *NICHOMACHEAN ETHICS***

CHAPTER

21

REFLECT

Any successful problem solver develops an array of techniques to spark her creative process. Bryan Mattimore describes dozens of such techniques in his book *99% Inspiration*, every single one of which involves reflection on the nature of the problem and the possible solutions.

This reflection can be conscious, unconscious, intellectual, intuitive, planned, unplanned, or all of the above. The key is to give yourself time to reflect on the problem and let a solution bubble to the top of your consciousness.

Reading books like Mattimore's can be helpful, but each person must develop techniques that mesh best with her own problem-solving style. When I am faced with a particularly thorny issue, I go for a walk or a run, which stirs my creative juices and removes me from other distractions. You should do what works for you: many people listen to music, make pottery, take a bath, or clean house to attain the focus they need for true reflection.

I begin my reflections by listing what I know about the problem and what a satisfactory solution would entail (without knowing the exact solution you can still define the kind of information that would solve the problem at hand). I then try alternative approaches in my mind, analyzing each one in terms of likely effort expended and probability of success, considering the available data. I let my mind wander as I walk, which often leads to important insights. By the end of my walk (assuming I haven't been hit by a car during my reveries) I have at least one possible way to attack the problem.

It is a rare person who can think creatively in the face of constant interruptions. I can't do it and I suspect most others can't either. Put aside some time every day to reflect a bit. Turn off the phone, close the door, and put up a "Do not disturb" sign. Schedule this reflection period during your particularly productive times of day.

Brenda Ueland, in her book *If You Want to Write*, asserts that creativity only bursts forth when we are idle and thinking on our own, uninterrupted. The con-

stant busyness we often experience in modern life just gets in the way of our own creativity. That's why she advocates at least half an hour per day of this sort of idleness.

Joe Hyams, in *Zen in the Martial Arts*, reports how one martial artist (Bronislaw Kaper) takes this idea to its logical extreme and consciously schedules entire days to do nothing. By taking time to reflect, you too can use such pauses to your advantage.

Music is the space between the notes. — **CLAUDE DEBUSSY**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

22

GET UNSTUCK

Sometimes, try as I might, I just can't push past a seemingly insurmountable obstacle. The problem itself may be especially difficult, or it may pose personal challenges. Any successful problem solver develops tricks for "getting unstuck" in the face of such obstacles.

One of my most successful tricks is to break my normal routine. Some of my best insights have come on airplanes where I have time to reflect and have few of my normal distractions. Go for a walk or a run, ride your bike, or go on a trip. Call a friend you haven't seen in a while. Do something out of the ordinary and forget about the problem for a few hours. Brilliant insights often occur when you're relaxed and not thinking about the problem at all.

You can also try working on a related problem or one that's entirely different. Either way, you'll be able to start fresh with the more vexing problem after you've attacked something a bit more tractable.

An artist I met once described how, once he had an idea for a painting, he would start on the canvas immediately. Once he got "inside the picture," the difficulties would often just melt away. The same technique works for other kinds of problem solving. If a problem remains daunting, ask if there's some small question that you can answer (jump "into the picture"). Answering that question often results in some obvious follow-up questions that can also be answered, and addressing these usually puts you on the road to a solution.

A time-tested technique for breaking logjams is to describe the problem to someone else. One friend recalled for me how her mother helped her solve math problems in junior high school. "Mom couldn't remember the details about specific math techniques but suggested that I explain the problem to her. By the time I had given her the details I virtually always had a new insight for a solution. Explaining the problem forced me to think clearly."

The converse of this approach is also true, which is that clear and precise explanations of results indicate clear and precise thinking on the part of an ana-

lyst. Those who can't explain their own analysis in an understandable way are not clear thinkers, and you should steer clear of their work.

There is a series of questions that flow through my mind whenever I'm faced with a problem unlike any I've seen before. These questions, largely based on G. Polya's book (*How to Solve It*), are as follows:

- "What is the unknown" (i.e., the number I am trying to calculate)?
- What data are given, and which can be estimated using knowledge I already have? Are the given data correct based on other knowledge I have? Are there any additional data I can bring to bear on the problem?
- Can I restate the problem in different words?
- Can I draw a picture or graph framing the problem?
- Is there another similar problem I have already solved that could give me the key to a solution? Can I solve a related problem?
- Can I make some simple assumptions that might help me solve the problem more easily? Can I solve just one part of the problem? Can I solve a simplified version of the problem, using round numbers that are roughly right?
- Are there any answers that I know are wrong? (Sometimes you can get closer to an answer by eliminating those that cannot be correct.)

24

The study of problem solving, which Polya calls "heuristic," is still a nascent art. Each obstacle is an opportunity to create and refine your own heuristic process. Examine your problem solving style, and create techniques for getting unstuck that work for you.

EXERCISE

Ask people whose judgment you respect how they get unstuck; then, try their techniques. Were you surprised at any of their answers?

A mistake is an event that has not yet been fully turned to your advantage. — **EDWIN LAND**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

23

INQUIRE

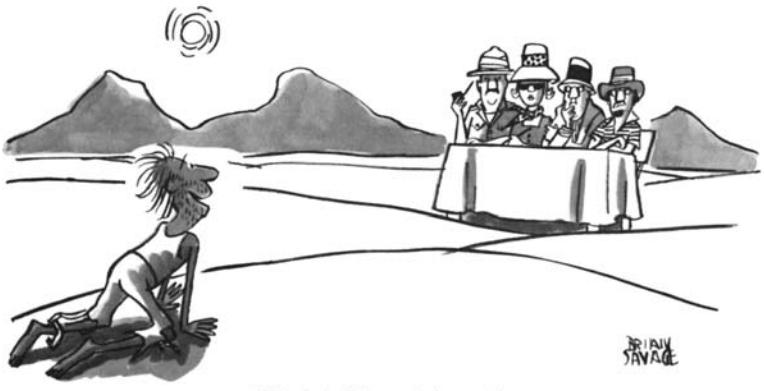
Quick—your boss just assigned you a task that is outside your area of expertise. What do you do?

I am often faced with this challenge, as are most analysts in business, academia, and government. Because the problem is new it's easy to have a "Beginner's Mind." My first move is always to pick up the phone. I have friends with a wide range of interests and usually there is at least one with direct knowledge of a particular topic. By asking these people focused questions I can quickly determine what is possible within the time allotted. I often develop these questions as part of my effort to "Get Unstuck."

I

22

Don't be afraid to approach experts with questions. They are often delighted that someone is taking an intelligent interest in their work and will usually bend over backwards to explain why their results are useful and important. The reason many people become experts is because of a passionate interest or concern. They *care* about their topic and want other people to care too. So go ahead and contact them. Your interest will make them happy, and you might just learn something along the way!



"Thank God! A panel of experts!"

© The New Yorker Collection 1978 Brian Savage from cartoonbank.com. All Rights Reserved.

If you don't know anyone who knows about the topic, try the Internet. Use one of the many search engines available for this task (try several, because they are all virtually guaranteed to give different results). Some search engines (e.g., <http://www.ask.com>) allow you to ask questions in plain English and receive focused results that number ten or twenty links (instead of the thousands you normally get from such searches). To hone your Internet research skills, consult some of the recent work that teaches these skills (e.g., <http://dir.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web>).

As discussed earlier, an important tip in Internet research is to consult the Frequently Asked Question sites (FAQs) that exist on the Internet for a truly amazing number of topics <http://www.faqs.org>. FAQs are compiled by experts and are often a superb place to begin your research. Use one of the many excellent Internet search engines to find FAQs that are relevant to your inquiry.

The web site <http://www.Amazon.com> is a surprisingly good place to do research. The main purpose of this site is to sell books, but its search routines are easy to use, and the site presents for each book a list of other books commonly purchased by people who bought the featured book. It also allows you to see other books by the same author with one click of the mouse. There are millions of books in the Amazon.com database, which is more than enough to cover most topics thoroughly. Even out-of-print books are included. I spent an instructive couple of hours exploring the site early one morning, and I came away impressed.

Another place to begin is a major reference work. For example, you can find something related to virtually any query in *The Statistical Abstract of the U.S.* or in an encyclopedia. The general information contained in these works can often be enough to get you started untangling even the thorniest research problem. For many social and physical science topics, check out the many different *Annual Reviews* that summarize the latest work in these fields.

If you are still stymied, talk to the reference librarian at the local library. Librarians have an uncanny knack for tackling just this sort of inquiry, and they are highly skilled professionals. They also love the challenge of finding new and interesting information. At a minimum, the librarian should be able to point you to the standard reference work in any area of knowledge as well as to the relevant trade journals (which can be treasure troves of information about real-world markets and technologies).

Librarians play a crucial role in research. The importance of these professionals came to the fore in a reporter's tale of searching for information in the spanking new Main Library in the city of San Francisco.[53] The reporter tried searching for books on a probably mythical Christian king who was the impetus for several important missions of exploration during medieval times. Although his topic may seem obscure, the reporter's experience will ring familiar to most. He tried first to use the library's electronic card catalog and came up with two listings on his subject, one of which seemed irrelevant. On the advice of the reference librarian, he then logged onto the library's on-line database and found seven new listings that didn't show up in the first search. Then he found another new listing by changing the order of his search terms from "Prester John" to "John, Prester." Finally, he consulted the old card catalog that the computer terminals had largely displaced and found an additional reference (a book that was so old that it hadn't been entered into the computerized database and was kept not in the main collection but in a storage annex).

Without the help of a librarian, this reporter might have missed some key sources. Only a librarian could tell him the tricks that revealed the sources he sought in an expeditious manner. Trial and error can yield results, but your search will be far quicker if you consult a librarian first.

Working outside your area of expertise is difficult, but can be rewarding. As we know from the history of science, some of the most important discoveries were made by scientists formally trained in another discipline. Developing the skills to explore new fields will serve you well in any endeavor you undertake.

Minds think with ideas, not information. No amount of data, bandwidth, or processing power can substitute for inspired thought.

— **CLIFFORD STOLL**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

24

BE A DETECTIVE

One of the defining features of the literary character Sherlock Holmes is his scientific approach to solving mysteries. He has an immense store of knowledge of past criminal activity upon which he can rely. He systematically collects information from the scene of the crime and notes seemingly trivial details that might bear upon his case. He does not prejudge the case in advance of the facts. Finally, he is skilled in weaving alternative theories that might fit the facts.

Holmes' approach is not unlike that required to solve problems in the modern world. Any problem solver worth her salt knows as much as possible about data and previous work on the problem at hand (if you're new to it, then go the library and read). Attention to detail is also essential, as is not allowing your initial snap judgments to color your interpretation of parts of the puzzle. And weaving alternative theories to fit the facts is perhaps the most important skill of a seasoned problem solver.

Spreadsheets[54] are my favorite problem-solving tool because they allow me to
26 create many different scenario calculations easily and quickly. Business people use them to create financial projections for their business plans. Scientists use them to analyze data from their experiments. Policy analysts use them to project the effects of proposed government policies on people's lives and on the economy as a whole. In each case, spreadsheets enable both quick scoping calculations and extremely detailed analysis.

Sometimes in creating a spreadsheet you discover an intermediate or final result that is either internally inconsistent with another such result from the same calculations or that cannot be right based on first principles. An example of the first kind of inconsistency would be to have a business plan that showed operating losses for the first five years of the forecast but that showed no expenditures of startup capital after five years (the company would need to use the startup capital to pay for the operating losses until the company became profitable). An example of the latter inconsistency would be to project that a small

startup company would gain 99% market share within a few years and attract no competitors (any time an important business niche appears, competitors almost always pop up to contest it).

When you discover such inconsistencies in your spreadsheets, Holmes' technique of systematically investigating alternative theories comes into play. Begin by making a list of possible reasons to explain why the calculation is in error; then, trace the flow of data through the calculations, starting from the point at which you noticed the inconsistency. Look at each and every input to a particular calculation to see whether you've made any obvious errors in entering formulas or data; trace the calculations all the way back to the beginning. If, as is commonly the case with complicated calculations, you still can't figure out what's wrong, recreate your calculation in a new spreadsheet, usually using a simplified version of the calculation and plugging in data that are roughly right (and that are nice round numbers like 100 or 1,000). This approach allows you to begin with a clean slate. Often when you are immersed in calculations you can forget small details. Starting fresh in this way helps you catch those errors.

One quick check you can do on your spreadsheet calculations is to vary a key input number, and see if the results change as you expect. If you increase that number by 50%, do you expect the result to increase or decrease by a comparable amount, or by half that amount? If you make that number close to zero or extremely large, do the results behave as anticipated? If your results vary in an unexpected way or in unusual proportions, your calculation is wrong or your conceptualization of the problem is flawed in some way. Either way, you've learned something.

If two complex calculations should yield the same result but do not, break down each calculation into its component parts and compare each part of the first calculation to the corresponding one in the second. This technique will make it obvious where in the calculations any differences lay. For example, assume that a formula in a spreadsheet has three terms that add to a total (Total = A + B + C) and that an independent calculation has three terms that should also add to that total (Total = X + Y + Z). Assume also that the value of each term in the first calculation should correspond to the value of each term in the second calculation (A should equal X, for example). If the result of the two calculations is different, then compare A to X, B to Y, C to Z, and see if the differences in the overall total can be explained by differences between any of the individual components. You can usually narrow down the discrepancy to one

or more of the terms in the equation and eliminate some possibilities. Although tedious, this technique can help you identify the potential explanations for such discrepancies. It is especially useful when each of the terms is, by itself, a complex calculation.

When working with spreadsheets, many people copy and paste data from one sheet into another. This practice is a pernicious one because it makes updating your calculations later a difficult task and makes it nearly impossible to determine how the calculations flow once they are no longer fresh in your mind. Obey this golden rule: *Never paste values from one spreadsheet into another*. Link spreadsheets together directly so that when you change your calculations in the first spreadsheet, the other one is automatically updated. Following this rule helps you track down errors in the calculation much more quickly. It also saves you time if you have to change input assumptions later, and it creates a special kind of documentation of your calculations.

33

If your results look too good to be true, they probably are. Be suspicious of your own calculations when you first create them. *Dig into the numbers* to determine whether you can trust them.

27

Another common technique used by detectives is to consider the motives or interests of the person or institution putting forth a particular viewpoint. I suspect, for example, that a lobbyist for the tobacco industry might be prone to overlook evidence that smoking is harmful. Seek out sources whose interests coincide with the truth, and discount somewhat the opinions of those with a vested interest. However, remember that having an interest or an opinion is not evidence of being biased any more than *not* having an opinion is evidence of being unbiased: some people (a rare few) are honest even to their own detriment.

14

When you have eliminated the impossible, whatever remains, however improbable, must be the truth. — **SHERLOCK HOLMES**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

25

CREATE CONSISTENT COMPARISONS

In the real world, so many things change at once that it is often difficult to decipher cause and effect. One of the most compelling techniques for making a point in the face of this complexity is to find or create a consistent comparison that eliminates the complicating factors and illustrates your point in an unambiguous way. A pure consistent comparison is one where the only feature differing between two examples is the one that illustrates your point. Even when there are some minor differences, you can often correct for these and still create a powerful comparison.

Isolating the root cause by creating an experiment is one of the most fundamental techniques of modern science. When medical researchers create "control groups" to compare to people being given an experimental drug, they are creating a consistent comparison or as near to one as is humanly possible. Although you can't always create such an experiment, you can often uncover one if you look carefully. I discovered such a case in the early 1990s, which led me to write one of my favorite papers. Since the 1970s, U.S. automakers had insisted that they were unable to increase the fuel efficiency of passenger cars in a cost-effective way. In 1992, Honda introduced a new Civic hatchback that was the same size and had the same engine power as the model it replaced, but it was 56% more fuel efficient than the previous year's model. I read news articles about this new car and badgered various friends who work on transportation issues to investigate this case further. After several months of fruitless nagging I decided to explore it myself, in collaboration with two colleagues from the Union of Concerned Scientists.

The two cars differed slightly in features but were essentially identical except for their fuel economy. The case was therefore a good one to use to estimate the actual retail costs of improving fuel economy in the real world. After correcting for the costs of the different features (based on dealer cost and list prices), the more efficient Civic (the 1992 VX) cost about $700 more than the

1991 Civic DX, and saved about 110 gallons of gasoline per year for typical drivers. The added cost would pay for itself in about five years at then prevailing fuel prices, making it a cost-effective investment.

The example was an exceptionally powerful one because the size and engine power of the two vehicles were identical, and the feature list was nearly so. As a check on the calculation of the retail cost, we also investigated the projected retail cost of the technologies used to improve the VX's fuel economy, based on
18 *independent engineering estimates* and found that these estimates were quite close to our calculation of the actual retail cost difference between the 1991 DX and the 1992 VX. This independent check gave us confidence that we were on the right track.

The 1993 Civic VX gets about fifty miles per gallon on the freeway yet holds four passengers comfortably. It's not a luxury vehicle by any means, but it's a testament to what automakers can do if they put their minds to it. The paper documenting this case subsequently won a prize from the National Research Council, which is an indication of the compelling nature of a well-chosen consistent comparison.

The cure for boredom is curiosity. There is no cure for curiosity.

— **DOROTHY PARKER**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

26

TELL A GOOD STORY

3

Life is about making choices in the face of uncertainty. One oft-neglected approach to such decision making is the art of building scenarios. The purpose of scenario analysis, as explained by Peter Schwartz in his now classic book *The Art of the Long View,* is to explore several possible futures in a systematic way. No matter how things turn out, the analyst will have on the shelf a scenario (or story) that resembles that future and will have thought through in advance the likely consequences of that scenario on the decisions at hand. Such a story "resonates in some ways with what [people] already know and leads them from that resonance to reperceive the world."

The steps for successful scenario development are as follows:

- *Define the question:* Scenarios should be developed in advance of some major strategic decision so that when the time to decide arrives, the options are laid out in detail and the implications of each decision path are made manifest. Such a decision might be 'Should our company expand into new markets next year, or should we focus on expanding our market share in our current markets?'
- *Determine the key factors affecting the system:* The next step is to identify the key factors determining whether the focal decision will succeed or fail. These factors might include such things as customer acceptance of a new technology or competitor response to the new strategy. They could also include demographics, inflation, commodity prices, technological change, public policies, or political instability. Some of them are beyond the control of any individual or institution ("predetermined elements"), and others are subject to human influence.
- *Evaluate factors by importance and uncertainty:* Of the trends and driving forces identified in the previous step, a small number will be both highly

important and highly uncertain (see also Morgan and Henrion's book *Uncertainty*). These are the issues upon which differences in the scenarios will be based. Predetermined elements don't differ among scenarios.

- *Choose the scenario set:* The most important and uncertain factors should be combined in various ways to define the set of scenarios to be considered. These factors are the dimensions of uncertainty that the scenarios will explore. For example, the scenario set for an analysis of the costs of reducing carbon emissions might include variations in fuel prices and costs of different technologies, which are two of the key sources of uncertainty. One scenario (a worst case for carbon reductions) might combine low fossil fuel prices with pessimistic technology costs for non-fossil sources. Another scenario might combine high fossil fuel prices with optimistic technology costs. The set of scenarios chosen also frames the story lines that will emerge from the scenarios.
- *Add detail to the scenarios:* Each of the key driving forces and factors from the second step should be assessed for each of the main scenarios. Usually, these factors can be assigned relative rankings (e.g., 'Inflation should be higher than average in scenario 1 and lower than average in scenario 2'), which can then be translated into actual numbers. The exact numbers are not so important as long as they reflect in a credible way people's expectations about how the future might unfold in a particular scenario. The scenario must then be woven into a plausible story describing a course of events that would result in the scenario's final outcome. Each scenario must be named in a way that vividly reminds people of its essence.
- *Determine the implications*: Revisit the decision identified in step 1, and assess how that decision would fare in each of the different scenarios. If the decision (e.g., expanding into new markets) would lead to success (as defined in step 2) for all or most of the scenarios considered, that decision is probably a wise one. If it is likely to be successful in only one scenario, the decision is "a high-risk gamble . . . especially if the company has little control over the likelihood of the required scenario coming to pass."
- *Define metrics for success:* Once the scenarios have been developed, it is crucial to identify a few key indicators by which decision makers can monitor the evolution of events. These indicators are used to determine which of the different scenarios best matches up with actual events and can be used to trigger an institutional response (e.g., 'if inflation gets up to 7%

> and the stock market goes down 15%, then do X'). If you can't measure it, you can't manage it, so defining appropriate indicators is crucial to success.

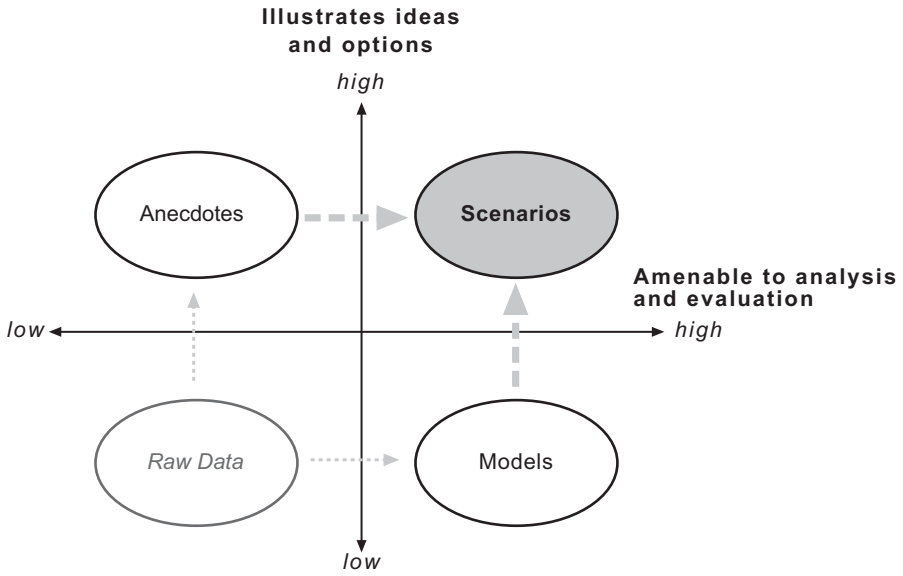
Schwartz believes that scenarios are successful when "they are both plausible and surprising; when they have the power to break old stereotypes; and when the makers assume ownership of them and put them to work." The institutional process of creating scenarios (what Schwartz calls "holding a strategic conversation") can also reveal new opportunities for an organization and promote team building among its employees.

Another name for scenario analysis is the commonly used scientific technique of creating a thought experiment. All scientists use this technique, in which they imagine certain real-world complexities to be absent and then think through the consequences for the situation they are investigating. For example, Galileo imagined a world in which the complicating effects of friction did not exist and, by doing so, was able to make great strides in what would come to be called the field of kinetics (the study of moving objects). Such thought experiments are as valuable in business as in science. Instead of limiting your imagination by current circumstances and constraints, imagine a world in which some key constraints don't exist, and ask yourself how you or your institution might respond.

Most of Schwartz's examples of scenario analysis have only a small quantitative component, but many other futurists are obsessed with numbers and computer models. Most explorations of the future with which I am directly familiar err by focusing too much on the mechanics of forecasting and quantitative analysis (e.g. on particular modeling tools) and far too little on careful scenario development based on the procedures described above. Quantitative analysis can lend coherence and credence to scenario exercises by elaborating on consequences of future events, but modeling tools should support that process rather than drive it, as is so often the case.

The relationship between data, anecdotes, models, and scenarios is summarized in Figure 26.1, adapted from Ghanadan and Koomey (2005). Raw data are difficult both to analyze and interpret, so people rely more on anecdotes (which illustrate concepts well but have little quantitative rigor) or models (the results of which often don't tell a coherent story but are at least easy to evaluate in a quantitative way). Good scenarios combine the narrative power of anecdotes with the additional quantitative rigor that comes from the use of models (even simple ones).

FIGURE 26.1. How scenarios relate to models, stories, and data



In my years of involvement with development of national energy policy, I have been most struck by how few resources are devoted to sensible scenario development and associated data and how many are devoted to the development of different modeling tools to assess such policies. Computer tools are sexy and appealing (at least to the funding agencies). Data and scenario analysis, upon which the results generally hinge, are virtually always given short shrift.

Tens of millions of dollars are spent every year on models whose capabilities are redundant with others, usually because a particular agency with money wants its own in-house modeling capability and is unwilling, because of institutional rivalry or personal biases, to adopt a pre-existing framework. Policy makers fail to realize that models are ALL unable to predict the future in an accurate way and that small improvements in modeling methodology are made irrelevant by inadequate scenario development.

The best way to predict the future is to invent it. — **ALAN KAY**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

27

DIG INTO THE NUMBERS

© 2001 by Tom Chen. All Rights Reserved.

Anyone who has delved into data from the real world knows that they're messy. Survey takers can write down the wrong response. People entering data into a computer can type in the wrong numbers. Computers sometimes garble data because of software bugs (especially when converting one file format to another). Data formats become obsolete as software changes. Electronic data

recorders break or go out of adjustment. Analysts mislabel units and make calculational errors.

"Cleaning the data" to correct for such problems is almost always necessary, but this step is often ignored. Always ask about the process used to clean the data. If the response is just a blank stare or worse, you know you're in trouble. Many companies (like AT&T and Sega of America) now assign people to check for data quality in the face of all the problems associated with real data.

Some of the more famous examples of how small mistakes in data processing can have disastrous results are found in the space program. The Mariner I spacecraft went off course soon after launch and had to be destroyed. The cause of the malfunction was a missing hyphen in a single line of Fortran code. Arthur C. Clarke later quipped that this $80 million launch was ruined by "the most expensive hyphen in history."[55] More recently, NASA's Mars Climate Orbiter was lost in space because engineers on the project forgot to convert English units to metric units in a key data file, a mistake that cost scientists years of work and taxpayers $125 million.[56]

Bad data can have a real cost for a business. If a 500,000-piece mailing uses an address list with an error rate of 20%, the company wastes $300,000 by sending mailings to incorrect addresses. It loses even more money because among those 100,000 missed prospects are about 1,500 people who would have become customers and purchased thousands of dollars worth of the company's products over their lifetime. The losses from these missed customers can be many times greater than the immediate direct losses.[57]

It is crucial to pore over raw survey data to check for anomalies before doing extensive analyses. For example, typographical errors can lead numbers to be ten, one hundred, or one million times bigger than they should be. Looking over the raw data can help you identify such problems before you waste time doing analysis using incorrect numbers.

Bad data can ruin your credibility and call your work into question. Even if there's only one small mistake, it makes your readers or listeners wonder how many other mistakes have crept into your analysis. It's difficult to restore your credibility after some obvious mistake is revealed, so avoid this problem in the first place. Dig into your numbers and root out these problems *before* you finalize your paper or talk.

SOME SPECIFIC ADVICE

If your data come from a non-electronic source, type the numbers into the spreadsheet yourself (assuming the amount of data is manageable). There's no substitute for this effort, even if you are a highly paid executive, because it will help you identify inconsistencies and problems with the data and give you ideas for how to interpret them. This technique also gives you a feeling for the data that can't be replicated in any other way. You will almost surely see patterns and gain unexpected insights from this effort.

Check that the main totals are the sum of the subtotals. Most documents are rife with typographical errors and incorrect calculations. You should therefore not rely blindly on any data source's summations but calculate them from the base data. You can check your typing accuracy by comparing the sums to those in the source of data. If they match exactly, it is unlikely that your typing is in error. Even if you don't check these sums, you can bet that some of your readers or listeners will. Do it yourself, and avoid that potential embarrassment.

Check that the information is current. Don't forget that business and government statistics are revised all the time. Make sure you know the vintage of the input data used in the analysis. For example, don't compare analysis results generated using one year's census data with those based on another year's data (unless your sole purpose is to analyze trends over time).

Check relationships between numbers that should be related in a predictable way. Such comparisons can teach valuable lessons. For example, when examining data on carbon emissions of different countries, a newcomer to the field of greenhouse gas emissions analysis might expect that the amount of carbon emitted per person would not differ much among industrialized countries. In examining such data for 2004, however, we find large differences in carbon emitted per person, from less than 1.6 metric tons/person/yr in Switzerland to about 7 metric tons/person/yr in Luxembourg. Determining why such differences exist is the logical next step, which will inevitably lead to further analysis and understanding.

Check that you can trace someone else's calculation in a logical way. If you can't, you can at least begin listing the questions you need to answer to start

tracing the calculation. Ultimately, if you can't reproduce the calculation, the author has broken a fundamental rule of good data presentation and his analysis is suspect.

Compare the numbers to something else with which you are familiar, as a "first-order" sanity check. These comparisons can show you whether or not you are on the right track. Presenting such comparisons in reports and talks can also increase your credibility with your readers or listeners because it shows that your results "pass the laugh test."

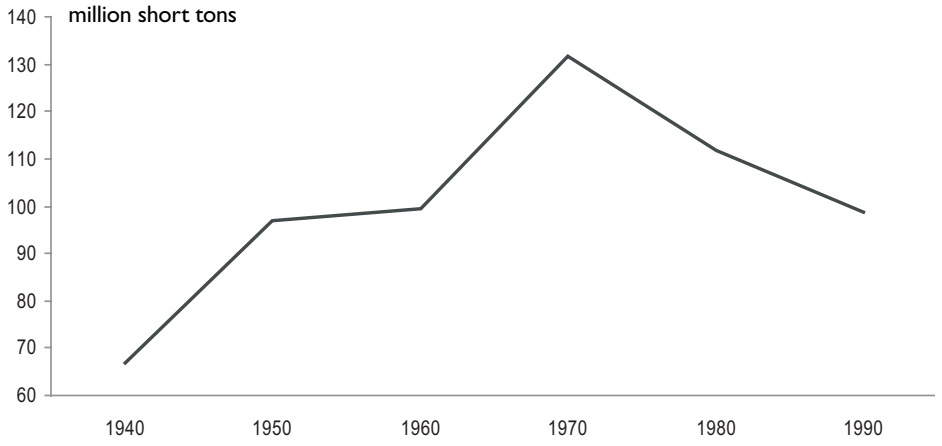
Normalize numbers to make comparisons easier. For example, the true size of total U.S. gross national product (GNP) in trillions of dollars per year is difficult to grasp for most people but if normalized to dollars per person per year will be a bit more understandable. Common bases for such normalizations are population (per person or per capita), economic activity (per dollar of GNP), or physical units of production (per kilowatt hour or per kilogram of steel produced).

If you have information that extends over time ("time series data"), normalize it to a base year to enhance comparisons. By expressing such data as an index (e.g., 1940 = 1.0), you can compare trends to those of other data that might be related. For example, if you plot U.S. raw steel production (**Figure 27.1**), population (**Figure 27.2**), and GNP (**Figure 27.3**) in separate graphs, it is difficult to gain perspective on how fast steel production is changing over time relative to these other two important determinants of economic and social activity. However, if you plot steel production over time as an index with 1940 = 1.0 (see **Figure 27.4**), you can plot population and GNP on the same graph. Such a graph will instantly show whether growth rates in the data differ.

In this example, real GNP grew by a factor of more than five from 1940 to 1990. U.S. steel production roughly doubled by 1970 and then declined by 1990 to roughly 50% above 1940 levels. The population in 1990 just about doubled from 1940 levels.

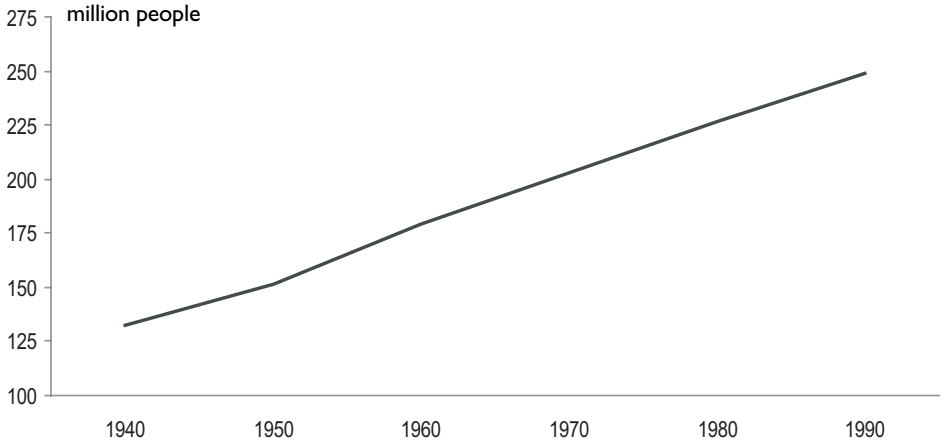
The trends for U.S. steel production in Figure 27.4 dramatically illustrate the changing fortunes of steel in a postindustrial economy. Just after World War II, steel use per capita increased as automobile ownership expanded and use of steel for bridges and other construction also increased. After 1970, the steel industry began to face serious competition from foreign steel makers as well as

FIGURE 27.1. U.S. production of raw steel 1940-1990 [58]



NOTE: Y-axis does not start at zero to better show the data.

FIGURE 27.2. U.S. population 1940-1990 [59]



NOTE: Y-axis does not start at zero to better show the data.

FIGURE 27.3. U.S. gross national product 1940-1990 [60]

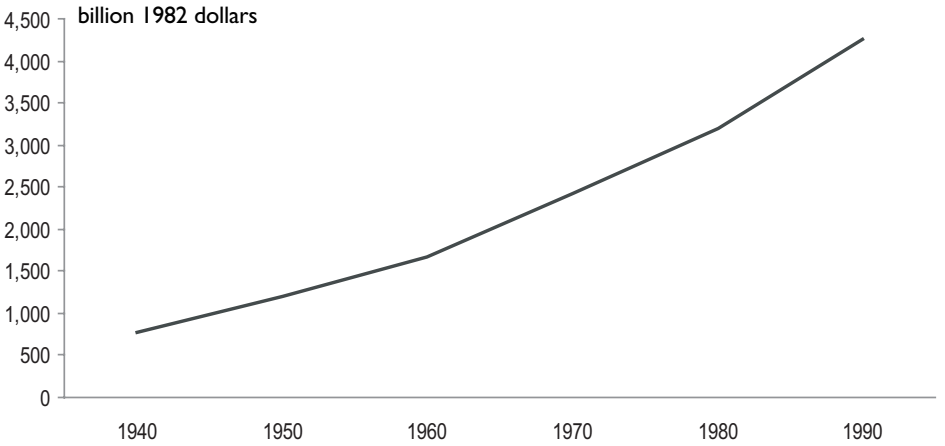
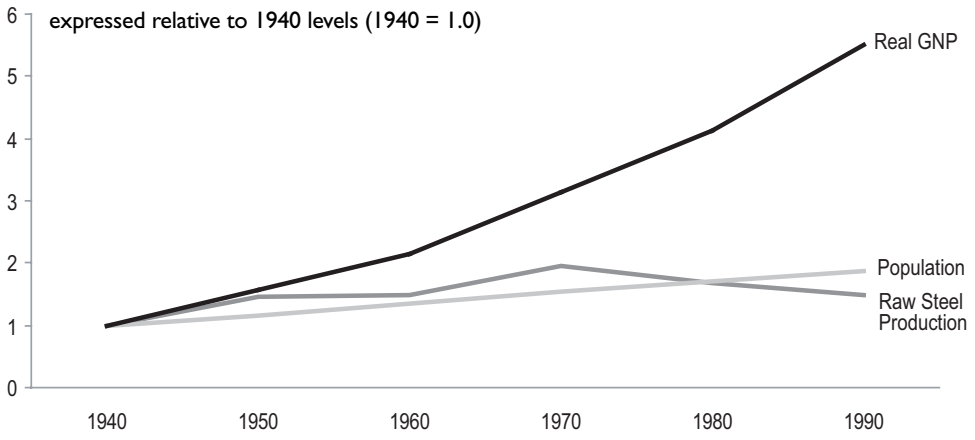


FIGURE 27.4. U.S. raw steel production, population, and gross national product 1940-1990



from alternative materials such as aluminum alloys and composites. Consequently, U.S. production declined even as the population increased and real GNP went through the roof.

Break problems into component parts. Explore analysis results by examining the factors that led to those results. For example, suppose someone tells you that the market capitalization of Google in November 2007 was about $226 billion and that of General Electric (GE) was $412 billion. What steps should you take to understand what these numbers mean?

Market capitalization is the product of the number of shares outstanding and the stock price per share, as shown in the following equation:

$$\text{Market Capitalization (\$)} = \#\ \text{Shares} \times \frac{\text{Stock Price}}{\text{Share}}$$

The stock price per share can be further broken down into the product of the earnings per share and the price-to-earnings ratio, yielding the following simple model:

28

$$\text{Market Capitalization (\$)} = \#\ \text{Shares} \times \frac{\text{Earnings}}{\text{Share}} \times \frac{\text{Price}}{\text{Earnings}}$$

The product of the number of shares and the earnings per share gives the total annual earnings (profits) for each company. Substituting in the previous equation, we get:

$$\text{Market Capitalization (\$)} = \text{Total Earnings} \times \frac{\text{Price}}{\text{Earnings}}$$

If we divide both sides of this equation by annual revenues, we get

$$\frac{\text{Market Capitalization}}{\text{Annual Revenues}} = \frac{\text{Total Earnings}}{\text{Annual Revenues}} \times \frac{\text{Price}}{\text{Earnings}}$$

28

All of these equations represent variations on the same model. Different forms of the model will be useful at different times. We chose the last version of the model for the discussion below because we can ascribe real meaning to the business factors affecting the two terms in that model.

For most companies, the basic information for the model is readily available on the Web, so that is the best place to start (you could also go to the library). **Table 27.1** summarizes the key financial parameters for calculating market capitalization for the two companies.

The first thing to notice about the financial statistics for these two companies is a huge disparity. While GE's market capitalization in November 2007 was about twice as large as Google's, GE's revenues were almost twelve times larger. All other things being equal, we might expect that companies with similar market valuations would also have similar revenue streams. All other things are *not* equal, however, and finding out why will help illustrate this important analytical technique.

If Google's market capitalization per dollar of revenues was the same as GE's, we would expect that its market capitalization would be only one-sixth as large as it is[61] and we need to explain this six-fold discrepancy. To do so, we examine the components of the last equation above. The first component is earnings per dollar of revenues, and the second component is the price-to-earnings ratio.

As Table 27.1 shows, Google's earnings as of November 2007, were about twice as big as GE's per dollar of revenues, which accounts for about a factor of two in our six-fold difference. The price-to-earnings ratio also differed between the two companies. Apparently, the stock market valued one dollar of

TABLE 27.1. Comparison of financial statistics for Google and General Electric (GE)

	Units	*GE*	*Google*	*Ratio Google /GE*
Annual revenues	B$ 2007	174	15.0	0.086
Number of shares	Billions	10	0.3	0.03
Price per share	$2007/share	40	722	18
Diluted Earnings per share (EPS)	$2007/share	2.2	12.8	5.9
Trailing price-to-earnings (PE) ratio	Ratio	19	56	3.0
Total earnings	B$ 2007	22	4.0	0.18
Earnings per dollar of revenues	$/dollar	$0.13	$0.26	2.1
Market capitalization	B$ 2007	412	226	0.55
Market cap. per dollar of revenues	$/dollar	$2.4	$15	6.4

(1) B signifies billions (thousand millions for British readers).
(2) PE is the share price to earnings ratio.
(3) Market capitalization (measured as an intraday value) is the product of # of shares and price per share. Price per share is the product of EPS and PE ratio.
(4) Values taken from Yahoo! Finance on November 4, 2007.

Google's earnings three times as much as one dollar of GE's earnings, which accounts for the remaining difference.

By breaking these numbers into their component parts, we have been able to isolate the two key reasons for the sixfold discrepancy identified above. The next step is to explain why these two reasons exist.

Google's greater profitability per dollar of revenue reflects the difference between industrial manufacturing and Internet software development. The latter has very low marginal costs of reproducing the product and high gross margins. Google's higher price-to-earnings ratio reflects the market's belief that its dominance in Internet technology will allow the company to continue to generate growth in earnings at a pace vastly exceeding that of traditional industrial enterprises.

Dissecting analysis results by comparing ratios of key parameters is a powerful approach, and one to use frequently. Any time you have two numbers to compare, this kind of "ratio analysis" can lead to important insights into the causes of underlying differences between the two numbers.

Creating a presentation or an executive summary for a complex report always involves boiling an analysis down to the essentials. First, create equa-

tions (like the ones above) that calculate key analysis results as the product of several inputs multiplied together. Then, determine which of these inputs affects the results most significantly. Creating such models can help you think systematically about your results. 28

APPLYING THESE LESSONS

Let's look at a specific example to make this advice more real. The web site <http://www.stateline.org> has time series data by state for a variety of items, including educational and environmental expenditures as well as tax revenues. As an exercise, I downloaded data from the site and incorporated it into an annotated Excel spreadsheet (statedata.xls) so that readers could explore the various steps I took to clean, analyze, and understand the data. You can download this spreadsheet at <http://www.numbersintoknowledge.com>. The data in this example include population, personal income per capita, state tax collected, state expenditures on environmental agency budgets, and state expenditures on education.

I first saved each stateline.org HTML table as text, and imported it into Excel.[62] These raw data then needed to be cleaned up a bit. I deleted blank rows, removed the Virgin Islands and Puerto Rico from the per capita income columns (because not all data series had values for these U.S. territories), moved the U.S. totals to the bottom, and checked each column of data to make sure that the order of the states was identical (sometimes there are slight differences in data order within such files; it's worth at least spot checking). I also scanned the data for numbers that were suspiciously small or large and found none. In addition, I added an "independent U.S. total" row, so I could add up the state data and make sure it matched against the U.S. totals reported in the data.

I then adjusted the data so that all aggregate dollar figures were expressed in millions of dollars. I reordered the columns to put population and personal income first because they are more fundamental data that could be used to normalize the other data in various ways. I calculated total personal income by multiplying per-capita income by population. I adjusted 1996 current dollars to 1997 dollars using the implicit GDP deflator. The 1996 education expenditures were not given in the stateline.org data, but 1997 expenditures were given along

TABLE 27.2. Cleaned data from stateline.org for selected states

State	*Population Millions*		*Per-capita personal income 1997$/person/year*		*Total state tax collections M1997$/year*		*State Environment agency budgets M1997$/year*		*State Education expenditures M1997$/year*	
	1996	1997	1996	1997	1996	1997	1996	1997	1996	1997
California	31.9	32.3	25,610	26,314	NA	61,667	677	686	35,498	35,546
New Mexico	1.7	1.7	18,981	19,298	NA	3,102	38	37	2,512	2,516
New York	18.1	18.1	29,555	30,250	NA	34,865	300	329	17,602	16,243
South Carolina	3.7	3.8	20,017	20,508	NA	5,381	169	227	3,898	3,903
Texas	19.1	19.4	22,761	23,707	NA	23,025	408	387	17,425	18,303
West Virginia	1.8	1.8	18,453	18,724	NA	2,906	235	196	2,114	2,386
Total U.S.	265.2	267.6	24,614	25,298	NA	443,493	7,020	7,481	268,423	275,821

(1) State tax collection data were not available for 1996.

with the percentage growth rate from 1996 to 1997. I then calculated 1996 education expenditures from these data. After all this "cleaning," my spreadsheet contained the data shown in **Table 27.2**, but for every state, not just for the six states shown here.

At this point, and only then, was I ready to analyze the data. I'd corrected for
33 inflation, and I'd checked to make sure that the data were not obviously flawed. I had also systematized my spreadsheet tables so that each of them started at the same row (for example, California would be on row 19 of all the spreadsheets). This technique makes it much easier to link to the spreadsheets and create calculations in an automatic way.

The first normalization was to calculate state tax collections, environmental agency budgets, and education expenditures on a per-capita basis (**Table 27.3**). On this basis, state environmental agency budgets are only tens of dollars per person per year; education budgets are generally about 50 times bigger per capita. Across the U.S., per-capita state tax collections amount to about 6.5% of total per capita personal income.

I also noticed that the education expenditures per capita were a very substantial fraction of total state tax collection per capita. In New Mexico, for example, these expenditures total more than 80% of all state taxes collected. This result seemed suspicious to me, so I investigated further.

9 I pulled out my *1997 Statistical Abstract of the U.S.,* and found Table 237 (School Expenditures, by Source of Funds in Constant [1993-94] Dollars: 1980 to 1994) and Table 477 (All Governments—Revenue, Expenditure, and Debt:

TABLE 27.3. Per-capita (PC) data for selected states

State	*Per-capita personal income 1997$/person/year*		*PC state tax collections 1997$/person/year*		*PC Environment agency budgets 1997$/person/year*		*PC Education expenditures 1997$/person/year*	
	1996	1997	1996	1997	1996	1997	1996	1997
California	25,610	26,314	NA	1,911	21	21	1,114	1,102
New Mexico	18,981	19,298	NA	1,793	22	21	1,468	1,454
New York	29,555	30,250	NA	1,922	17	18	971	896
South Carolina	20,017	20,508	NA	1,431	45	60	1,049	1,038
Texas	22,761	23,707	NA	1,184	21	20	913	942
West Virginia	18,453	18,724	NA	1,600	129	108	1,161	1,314
Total U.S.	24,614	25,298	NA	1,657	26	28	1,012	1,031

(1) State tax collection data were not available for 1996.

1980 to 1994). As usual, the *Statistical Abstract* tables give footnote references to the original sources, which would allow me to check the details if necessary.

At that time, the latest data from the *Statistical Abstract* were for 1994, but they were recent enough to serve as a check on the stateline.org data. Table 477 from the *Statistical Abstract* showed 1994 total state tax revenues of $373B in 1994$ ($396B in 1997$), and growth over the preceding four-year period of $70B, so I could easily imagine state tax collections reaching the $444B figure (in 1997$) from the stateline.org data by 1997. This calculation confirmed that the stateline.org data for tax collections were plausible. Taxes are not the only source of state revenue, however. Total state revenues include a category called "Current charges and Miscellaneous" in Table 477, and these additional revenues increase the revenues from state taxes by about 30%. To completely sort this out, I would have had to find the original source of these estimates and determine whether these "miscellaneous" revenues are used to support education in any way. I didn't search for this information in this case because of the illustrative nature of this example.

The education expenditures took a bit more investigation. One complexity is that some of the state expenditures on education come from money contributed by the Federal government (this money is not dependent on state tax receipts). Another complexity is that the state education expenditures include *local* government expenditures, but the state tax receipts do *not* include local tax revenues. Table 237 from the *Statistical Abstract* helped make this point crystal clear because the state expenditures on education in 1994 in this table total only

TABLE 27.4. Per-capita (PC) data normalized to the national average for selected states

State	*Per-capita personal income Natl average = 1.0*		*PC state tax collections Natl average = 1.0*		*PC Environment agency budgets Natl average = 1.0*		*PC Education expenditures Natl average = 1.0*	
	1996	1997	1996	1997	1996	1997	1996	1997
California	1.04	1.04	NA	1.15	0.80	0.76	1.10	1.07
New Mexico	0.77	0.76	NA	1.08	0.84	0.77	1.45	1.41
New York	1.20	1.20	NA	1.16	0.63	0.65	0.96	0.87
South Carolina	0.81	0.81	NA	0.86	1.71	2.16	1.04	1.01
Texas	0.92	0.94	NA	0.71	0.81	0.71	0.90	0.91
West Virginia	0.75	0.74	NA	0.97	4.88	3.85	1.15	1.28
Total U.S.	1.00	1.00	NA	1.00	1.00	1.00	1.00	1.00

(1) State tax collection data were not available for 1996.

TABLE 27.5. Per-capita (PC) tax collection and expenditure data as a percentage of per-capita personal income (PI) for selected states

State	*Per-capita personal income PC PI = 100%*		*PC state tax collections PC PI = 100%*		*PC Environment agency budgets PC PI = 100%*		*PC Education expenditures PC PI = 100%*	
	1996	1997	1996	1997	1996	1997	1996	1997
California	100%	100%	NA	7.3%	0.1%	0.1%	4.4%	4.2%
New Mexico	100%	100%	NA	9.3%	0.1%	0.1%	7.7%	7.5%
New York	100%	100%	NA	6.4%	0.1%	0.1%	3.3%	3.0%
South Carolina	100%	100%	NA	7.0%	0.2%	0.3%	5.2%	5.1%
Texas	100%	100%	NA	5.0%	0.1%	0.1%	4.0%	4.0%
West Virginia	100%	100%	NA	8.5%	0.7%	0.6%	6.3%	7.0%
Total U.S.	100%	100%	NA	6.6%	0.1%	0.1%	4.1%	4.1%

(1) State tax collection data were not available for 1996.

$165B in 1994$ ($175B in 1997$), which is significantly lower than the $276B shown in Table 27.2. The local expenditures must account for the rest.

If the local tax revenues are added to the state tax receipts, the fraction of tax receipts represented by education would be much lower than it appears from Table 27.3. We solved the puzzle, but it took comparison with an independent source to do it.

Another way to normalize the data is as a fraction of the national average, as shown in **Table 27.4**. This approach shows quickly which states are higher or lower than the mean. In this sample, California and New York have per-capita personal incomes above the national average (their state tax collections per

TABLE 27.6. Percentage growth 1996–97 in total and per-capita data for selected states

		Total				*Per Capita*			
State	Population	Personal Income	State Tax collections	Environmental agency budgets	Education expenditures	Personal income	State tax collections	Environment agency budgets	Education expenditures
California	1.3%	4.1%	NA	1.4%	0.1%	2.7%	NA	0.1%	-1.1%
New Mexico	1.1%	2.8%	NA	-2.2%	0.1%	1.7%	NA	-3.3%	-0.9%
New York	0.0%	2.4%	NA	9.6%	-7.7%	2.4%	NA	9.6%	-7.7%
South Carolina	1.2%	3.7%	NA	34.9%	0.1%	2.5%	NA	33.3%	-1.0%
Texas	1.8%	6.1%	NA	-5.2%	5.0%	4.2%	NA	-6.9%	3.2%
West Virginia	-0.3%	1.2%	NA	-16.8%	12.9%	1.5%	NA	-16.6%	13.2%
Total U.S.	0.9%	3.7%	NA	6.6%	2.8%	2.8%	NA	5.6%	1.8%

(1) State tax collection data were not available for 1996.

capita, as well as those of New Mexico, are also above the norm). South Carolina and West Virginia both show per-capita environmental agency budgets well above the U.S. average. In fact, these numbers are so much higher than the average that they probably warrant further investigation.

Per-capita state education expenditures in California, New Mexico, South Carolina, and West Virginia also exceed the national average, but, as we saw above, care must be used in interpreting these numbers. The federal government also contributes funds to education, and these funds are not always distributed equally on a per capita basis among the states.

Still another way to normalize the data is as a percentage of per-capita personal income, as shown in **Table 27.5**. This view of the data shows how important each of these items is compared to an important measure of the total wealth of our society.

Finally, I calculated percentage growth from 1996 to 1997 in total and per-capita data, as shown in Table 27.6. For time-series data over a longer time period, it is also useful to calculate an index relative to the base year, as shown in Figure 27.4 (above). New York and West Virginia show comparable growth rates between total and per capita data; the other states and the U.S. average show lower growth rates (or faster declines) on a per-capita basis than for the aggregate totals. Of course, two sequential years of time series data are not terribly illuminating, but the same techniques can be applied when analyzing longer time series.

CONCLUSIONS

When John Holdren was a professor at UC Berkeley, he taught a class called "Tricks of the Trade," in which he listed key pitfalls to avoid in data acquisition and handling. I have aggregated them below into four golden rules:

33

- *Avoid information that is mislabeled, ambiguous, badly documented, or otherwise of unclear pedigree*. Ambiguity and poor documentation are an indication that the quality control for such data is uneven at best and appalling at worst. Dig into the numbers a bit and find out whether it's carelessness or incompetence; make sure you believe the numbers before using them.
- *Discard unreliable information that is invented, cooked, or incompetently created*. If you find major inconsistencies, conceptual flaws, and omissions in the data, you'll need to discard that information, no matter how much it might help your analysis.
- *Beware of illusory precision.* Don't represent or interpret information as more accurate than it is. Carefully characterize uncertainty and variability in your data, and insist that others do so with their own.
- *Avoid spurious comparability*. Beware of numbers that are ostensibly comparable but are actually inconsistent. Create and use only consistent comparisons.

25

Holdren's advice when dealing with data is "Be suspicious, skeptical, and cynical. Assume nothing." Though it may sound paranoid to the uninitiated, such caution is an absolute necessity for the seasoned problem solver.

For every foolproof system, there is a bigger fool.

— **EDWARD TELLER**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

28

MAKE A MODEL

In this book, the term "model" refers to any conceptual (usually mathematical) model that can be used to understand the world. Models can range in scope from the back of an envelope to millions of lines of computer code. Anthony Starfield et al., in their book *How to Model It*, summarize nicely when they state "An explicit model is a laboratory for the imagination." Models allow you to experiment, test, and learn about the world in a systematic way without leaving your office.

SIMPLE MODELS AND BURIED ASSUMPTIONS

A model is an equation or set of equations that describe how something you care about changes when other things change. Below is a useful model that allows you to estimate gasoline consumption as a function of how many miles you drive and the efficiency of your car:

$$\text{Gas used per week (gallons)} = \frac{\text{Miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{Miles}}$$

You can estimate your car's average fuel efficiency based on past experience in a typical week. If you drive 100 miles per week, and your car consumes five gallons of gasoline in that period, your car gets 20 miles per gallon (mpg) on average. For every additional 100 miles you drive, you will consume another five gallons of gasoline.

This model is perfectly adequate for many purposes, but a more detailed model may be required in some cases. Say that you want to estimate your gas consumption for a special week-long, 1,000 mile trip that will involve much

more highway driving than would your normal pattern. The following model would then be useful:

$$\text{Gas used} = \frac{\text{City miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{City miles}} + \frac{\text{Highway miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{Highway miles}}$$

Let's say that the car's efficiency is 15 mpg in the city and 25 mpg on the highway, and that your big trip is 95% highway driving, as shown in **Table 28.1** (about 67% of the driving is on the highway in the typical driving pattern embodied in the simpler model). The more complex model indicates that the vehicle would consume a total of 41.3 gallons, while the simpler model (which is based on your typical driving pattern) yields an estimate of 50 gallons when applied to your special trip. The simpler model therefore overestimates gasoline use by 21% in this case.

There are many other subtleties that could be incorporated into such a model, which would be introduced as additional terms in the equation. Neither one of these models is precisely correct in a scientific sense although the second one is more generally applicable. Either can be appropriate in certain situations.

The assumption about how much city and highway driving you do is *buried* in the first model. It's easy to see how someone not aware of the difference in automobile efficiency in city and highway driving could misuse the first model and calculate an erroneous result. This lesson is a more general one about models: assumptions are often buried and can lead you astray.

TABLE 28.1. Simpler and more complex models of automobile fuel use

	Driving Mode	*Simpler model*	*More complex model*	*Ratio of Simpler to More complex*
Miles per gallon	City		15	
	Highway		25	
	Average	20	24.2	0.83
Miles driven	City		50	
	Highway		950	
	Total	1000	1000	1.00
Estimated gallons consumed	City		3.3	
	Highway		38	
	Total	50	41.3	1.21

SIMPLIFY, SIMPLIFY, SIMPLIFY

People are often intimidated by the use of computer models to conduct an analysis, but never forget that these models depend on data and assumptions that are based in large part on analysts' judgment. It is a rare model that does not contain dozens or hundreds of such assumptions; in fact, these data and assumptions determine the outcome in many cases. This heavy reliance on assumptions for parameters that are unknown is most prevalent in models used in business and public policy, but even models of the physical world can rely on assumptions that influence or determine the outcome.

In his real estate practice, Robert J. Ringer encountered sellers with fantastic expectations for the investment value of their buildings:

> As a general but very consistent rule, owners are so unrealistic when it comes to vacancy factors, replacement costs, and other expense items that are not readily ascertainable that I normally considered their projections and operating statements to be virtually meaningless.[63]

Using his electronic calculator and his knowledge of real estate, Ringer was able to determine quickly whether the building was salable at or near the owner's asking price. His simple conceptual model allowed him to create a quick estimate of net cash flow on the proverbial back of an envelope. This screening technique allowed him to avoid wasting time on deals that would never close because of the huge gap between the reality of the project's net cash flow and the fantasy of the owner's expectations.

The technique of using a simple and transparent model to evaluate the results of a more complex calculation is immensely powerful. Analysts at the International Institute for Applied Systems Analysis (IIASA) found this out to their chagrin when Will Keepin, a Visiting Scholar at IIASA, was able to almost exactly reproduce the results of a multiyear, multimillion dollar study using some of the study's key input assumptions and a hand calculator. Keepin showed that the study's results followed directly from the input assumptions. He concluded that the study's projections of future energy supply "are opinion,

rather than credible scientific analysis, and they therefore cannot be relied upon by policy makers seeking a genuine understanding of the energy choices for tomorrow."

Beware of big complicated models and the results they produce. Generally they involve so much work to keep them current that not enough time is spent
26 on data compilation and scenario analysis. Morgan and Henrion, in their book *Uncertainty*, devoted an entire chapter to such models and began by summarizing this fundamental truth:

> There are some models, especially some science and engineering models, that are large or complex because they need to be. But many more are large or complex because their authors gave too little thought to why and how they were being built and how they would be used.[64]

Such large models are essential only for the most complex and esoteric analyses, and a simpler model will usually serve as well (and be more understandable to your intended audience).

The importance of transparency cannot be overestimated. A model that your audience can actually grasp is inherently more persuasive than a "black box" that no one outside of a small circle of analysts understands. Transparent mod-
33 els for which the input data and assumptions are also well documented are even more compelling, but are, sadly, all too rare.

Don't be too impressed by a model's complexity. Instead, ask about the data
26 and assumptions used to create scenarios. Focus on the coherence of the scenarios and their relevance to your decisions, and ignore the marketing doublespeak of those whose obsession with tools outweighs their concern with useful results. Sadly, many modelers fall into this latter camp, and you would do well to avoid their work.

Programming today is a race between software engineers trying to build bigger and better idiot-proof programs, and the universe trying to produce bigger and better idiots. So far, the universe is winning.

— **RICH COOK**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

29

REUSE OLD ENVELOPES

Ross and Ross, Mattimore, and von Baeyer describe how the physicist Enrico Fermi used to dole out exam problems but not supply all the information necessary for a solution. His students (my former professor Art Rosenfeld was among them) were expected to make educated guesses about the missing parameters to solve these so-called "Fermi problems." This technique is not often taught in school, but it is one well worth learning. For virtually any problem, it is possible to create an approximate solution using information you know or can estimate from daily life experience. Most people don't believe this until they try it a few times, but it's true.

FROM OIL TO WATER

I read an article once that described how Global Water Corp. of Vancouver had signed a contract with an Alaskan city (Sitka) that would allow the company to harvest water from glacial runoff.[65] The company was going to take converted oil tankers and ship the water to China, where it would then be bottled and sold. I initially thought that this venture was absurd, but a few back-of-the-envelope calculations made it clear that there was method to this apparent madness.

The company signed a fifteen-year contract covering 4.8 billion gallons of water a year. The article claimed that half-liter bottles of water from Europe currently cost Chinese consumers about $5.25 while this new source of water would cost about $0.70 per half liter bottle ($1.40 per liter).

The oil tanker was originally meant to carry crude oil, which, at the time, was selling for about $20 per barrel. There are 42 gallons per barrel of oil, yielding a cost per gallon of about $0.50 (gasoline typically sold for twice that amount in the U.S. at that time because of the additional refining and trans-

portation costs to bring the fuel to gas stations, as well as federal and state taxes). There are about 4 liters per gallon (a liter is a bit more than a quart), so the price of oil per liter was about $0.12.

The bottled water would (assuming the article was correct) sell for $1.40 per liter, as described above, which is about ten times greater than the price of the original cargo for the tanker (oil). This calculation showed that the company could make some serious money shipping water to China, assuming that the market for bottled water was large enough to sell the tanker's contents. Of course, there are costs of cleaning and converting the tanker to hold pure water, and costs of packaging the bottled water that a gasoline vendor would not incur, but the costs of transport would be comparable. The one-time cost of converting the tanker would be relatively small when amortized over many trips. I guessed that the costs of packaging and marketing the water would at most amount to between ten and fifty cents per liter. These costs would reduce the profits of the venture, but it still seemed likely to be profitable after these adjustments.[66]

MORE CALCULATIONS

I once sat outside an air-conditioned store in Cambridge, Massachusetts that had its door propped open on a very hot day. Using some simple assumptions and conversion factors I had memorized, I calculated how many dollars per hour the store was wasting by refrigerating the outside world. That calculation was enough to convince the store manager to close the door (he probably never had anyone badger him about *that* before, even with all the crazy scientists wandering around Cambridge!).

Another calculation I did about ten years ago was to figure out the wattage of a candle. All I needed to assume was that wax is similar to petroleum (I had the energy content of petroleum memorized) and note the time it took for the candle to burn away. I also needed to know some conversions to turn English units into metric units. It turns out that typical dinner candles dissipate energy at a rate comparable to that of a 60 to 100 watt light bulb (of course, candles are much less efficient in creating light than are electric lamps, so they don't create nearly as much light as a bulb).

USING THE BACKS OF ENVELOPES

Such "back-of-the-envelope" calculations, supplemented by what Mattimore calls "smart guessing," are not just for scientists. A business plan for a new venture, for example, usually involves dozens of such smart guesses about sales of a product, costs of sales, overhead, and other parameters. The first draft of such a plan is essentially a back-of-the-envelope calculation that will later be refined and improved in discussions with experienced business people.

People involved in fast-moving new ventures usually don't have time to research issues thoroughly and don't have money to hire someone to do it for them (sometimes even the people they hire don't have time either!). Often, back-of-the-envelope calculations are the best that can be done in the time allotted. The law of averages keeps the accuracy of such calculations from deteriorating too much because you are just as likely to overestimate as to underestimate (assuming that you're not biased one way or another). Across all the assumptions you make, these deviations average out.

EXERCISE

Estimate how many cobblers there are in the U.S., using just information you have in your head. Check your answer and method against that used in the first chapter of John Harte's book *Consider a Spherical Cow*.

THINKING ABOUT THE PROCESS

There are many ways to create these quick calculations, but the first step is always to create a one-equation *model* by breaking the problem into its component parts. This equation might look like the relationship between market capitalization and earnings per share in the chapter titled "Dig into the numbers": 28 27

27

$$\text{Market Capitalization (\$)} = \#\,\text{Shares} \times \frac{\text{Earnings}}{\text{Share}} \times \frac{\text{Price}}{\text{Earnings}}$$

It might also look like the model of automobile fuel economy shown in the
28 chapter titled "Make a model":

$$\text{Gas used} = \frac{\text{City miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{City miles}} + \frac{\text{Highway miles driven}}{\text{Week}} \times \frac{\text{Gallon}}{\text{Highway miles}}$$

An important part of this process is balancing accuracy against the time needed to create the calculation. Back-of-the-envelope calculations are meant to
28 be quick, so they usually rely on relatively simple models.

Once the relationships between key inputs and the desired output are established, you'll need to determine which of the inputs you know and which you'll have to assume for purposes of the initial calculations. *Don't get hung up on a particular number.* The best thing to do is to carry out the calculation using the very roughest of approximations for those things you don't know, and refine the inputs later.

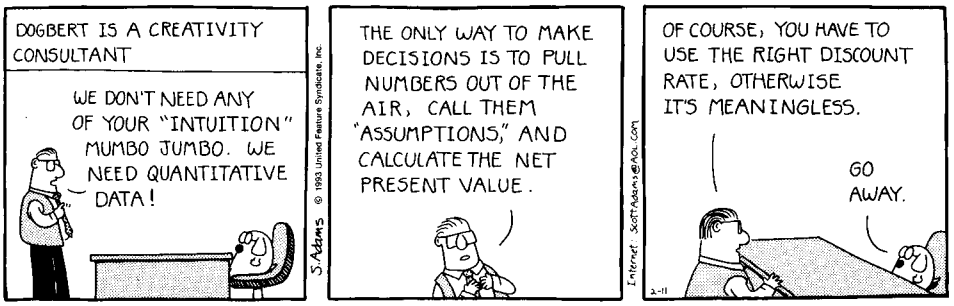
BOUNDING THE PROBLEM

When you've set up your calculation with your best guesses for the various input parameters, it's often instructive to identify those that are least certain and to determine a plausible range for them. You then carry through the calculation using the upper and lower ends of the ranges for all the inputs.

For example, if you only knew the fuel economy of your car very roughly, you might specify a range of 12 to 18 miles per gallon in the city and 22 to 30 miles per gallon on the highway for the "gas used" equation above. By combining the lower ends of both efficiency ranges in your calculation, you could determine an "upper bound" to how much gasoline you might use on the trip. This calculation could be useful in making sure that you don't ever run out of gas along the way or that you have enough money to pay for gas in your budget. Using the upper ends of the efficiency ranges would give you the minimum cost for gas that is likely for the trip.

26 These kinds of bounding calculations actually represent simple scenario analyses. If you compute ranges for the inputs and the results don't vary much from your best estimate, you've learned that the calculation is insensitive to the uncertainties you've identified, which is an important result. There are, of

The desire for quantification can go too far.


DOGBERT IS A CREATIVITY CONSULTANT
WE DON'T NEED ANY OF YOUR "INTUITION" MUMBO JUMBO. WE NEED QUANTITATIVE DATA!
THE ONLY WAY TO MAKE DECISIONS IS TO PULL NUMBERS OUT OF THE AIR, CALL THEM "ASSUMPTIONS," AND CALCULATE THE NET PRESENT VALUE.
OF COURSE, YOU HAVE TO USE THE RIGHT DISCOUNT RATE, OTHERWISE IT'S MEANINGLESS.
GO AWAY.
© 1993 United Feature Syndicate, Inc.
S. Adams
Internet: ScottAdams@AOL.COM

DILBERT: © Scott Adams/Dist. by United Feature Syndicate, Inc.

course, much more sophisticated ways of characterizing uncertainty, but for the purposes of the simple calculations discussed here, bounding the ranges for key parameters is often good enough.

APPROXIMATIONS

Most people think of numbers as immutable, but when doing rough calculations, it's perfectly OK to fiddle with the numbers a bit to make the calculations easier. For example, say you need to divide 3,500 by 40. Instead of reaching for your calculator, you can say "40 is only a bit more than 35, and 3,500/35 = 100. Therefore, 3,500/40 is a bit less than 100." These approximations can save you time when only a rough answer is needed. Don't be afraid to make these simplifications when you need to.

UNIT ANALYSIS

Another simple but powerful tool in back-of-the-envelope calculations is called "unit analysis." Think of the calculation we did above to determine the cost of oil in $/liter. When creating the model, you could just list the numbers and multiply them together, like this:

$$\text{Cost (\$/liter)} = \frac{20}{(42 \times 4)} = 0.12$$

Alternatively, you could list each of the terms with units attached so that it is obvious that they cancel out and yield the end result in units we desire:

$$\text{Cost (\$/liter)} = \frac{\$20}{\text{barrel}} \times \frac{1 \text{ barrel}}{42 \text{ gallons}} \times \frac{1 \text{ gallon}}{4 \text{ liters}} = \frac{\$0.12}{\text{liter}}$$

I initially scoffed when my high school chemistry teacher taught this methodical way of carrying through and checking our calculations. At the time, the calculations were so simple that I didn't need to be methodical and could just do them in my head. I've learned through experience, however, that calculations are rarely that simple in practice, and, even if they are simple, I still make mistakes because I'm tired or rushed. It pays to write out the units for your calculation to make sure you are calculating correctly.

OTHER HINTS

To become truly proficient at these techniques, memorize key conversion factors, at least in a rough way. Which ones you memorize depend on your chosen field. Energy and environmental conversion factors are most useful in my work, 9 but I have also memorized many other conversions because they often come in handy.

If you simply can't memorize such items, then create a small page with common conversion factors and keep it in your wallet. Graduate students at the Energy and Resources Group at UC Berkeley are routinely supplied with several of these pages during their course of study, and one of these is reprinted in 9 "Facts at your fingertips." While creating such a "cheat sheet" may brand you as a nerd in the eyes of your compadres, it will surely impress them when you can answer questions while only consulting your data sheet.

One modern tool that is invaluable for such calculations is the spreadsheet (proving that you don't need an envelope to create back-of-the-envelope calculations). Create quick calculations using a spreadsheet and plug in guesses for those parameters you don't know at the time. This approach allows you to think through the calculations without hindrance, and then go back later to check the input assumptions. Later modifications to these assumptions trickle through the calculations nicely without requiring much additional work.

CONCLUSIONS

Hans Christian von Baeyer points out that doing such calculations instead of relying on authority engenders self-confidence and independence. Achieving even occasional success in back-of-the-envelope estimation increases your inclination to tackle such problems in the future, thereby ensuring that you will gain further experience and self-assurance. To know in your heart that you can roughly estimate just about anything is a marvelous feeling of mastery.

EXERCISE

Find three quantitative assertions in today's newspaper, and do back-of-the-envelope calculations to check them. Did you find anything surprising?

Whether the problem concerns cooking, automobile repair, or personal relationships, there are two basic types of responses: the fainthearted turn to authority—to reference books, bosses, expert consultants, physicians, ministers—while the independent of mind delve into that private store of common sense and limited factual knowledge that everyone carries, make reasonable assumptions, and derive their own, admittedly approximate, solutions.

— HANS CHRISTIAN VON BAEYER

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

30

USE FORECASTS WITH CARE

The future is uncertain, but people keep trying to forecast it anyway. In an article with an ironically humorous and utterly extraneous subtitle—"Few Wall Street Analysts said that '96 would be this good: Forecasts for '97 may or may not be just as inaccurate"—Chet Currier of the Associated Press assessed the sorry record of most financial analysts in predicting the huge increase in stock prices in 1996.[67]

After describing how virtually all analysts predicted modest gains or little change from end-of-1995 levels, Currier states that the major market indices racked up gains of 20% or more in 1996. His conclusions about the value of such forecasts are particularly insightful:

> None of the errant prognostications cited above were wild guesses or interpretations of the leaf patterns in some hallucinogenic tea. They were the carefully considered projections of respected commentators who enjoy wide followings as financial analysts. All of them remain in their same jobs, or comparable positions of prominence, a year later, and are presumably busy right now preparing to make their projections for another year.
>
> As usual, these commentaries may well provide a useful point of departure for your own investment planning in 1997. *The reasoning behind a prediction is often the most interesting and useful thing about it* (italics added). But the experience of this year dramatizes how skeptically you may take all forecasts of something as unpredictable as the stock market—even from people who know what they are talking about.

Newspapers always trumpet the results of such forecasts, but, as the stock market example shows, you should be wary.

Steven Schnaars, in his book *Megamistakes: Forecasting and the Myth of Rapid Technological Change*, gives many examples of forecasts of the popularity of new products that were totally off the mark. In fact, about 80% of the

forecasts he examines were grossly inaccurate, which made him question the entire basis for predicting adoption of new technologies. He concluded, after reviewing dozens of examples, that explicit scenario analysis is far superior to most conventional ways of exploring the future. 26

Forecasts are used

- for planning (either institutional or personal);
- for advocacy (to lend credence or context to concern over a particular issue);
- for research (to assess the potential effects of current or expected future trends).

Most institutions create forecasts for their own budget planning. Such a forecast attempts to capture recent trends and allows the institution to plan for expected growth or declines in revenues. A typical forecast in this context would look one or two years ahead, sometimes three or four years in exceptional cases. Longer-term forecasts are used for government institutions (like Social Security) that have obligations spread over many years. Forecasts can also serve as part of a scenario-planning exercise.

An institution concerned with a particular issue will sometimes create a forecast that validates concern for that issue, as part of its effort to create free publicity. One edition of *The Sunday San Francisco Chronicle/Examiner* contained two separate examples where a forecast was cited for this purpose.[68] In one case, a California real estate investor and analyst predicted that housing prices in California would double over the next eight years. In another, the Cremation Association of America predicted that the rate of cremation, measured as the % of deaths that lead to cremation, would increase from 21% in 1997 to 38% in 2010. In both cases, a person or institution developed forecasts that showed why a particular issue of concern should also be of interest to the public.

The investor who predicted a doubling of California home prices was roundly criticized by his colleagues, but he succeeded in getting the attention of the press, in part because his prognostication conflicted so glaringly with the conventional wisdom. That strategy is a successful one that is widely used: predict something just a tad outrageous, and use your credibility from previous work to lend credence to your prediction; then talk to as many reporters as will listen in hopes they will mention you and your company in their columns. In

this particular case, the prediction was right on the mark—median California home prices more than doubled in nominal terms from 1997 to 2004 (see <http://www.realestateabc.com/graphs/calmedian.htm>).

More complex issues can require detailed forecasts to truly understand them, which usually involves computer modeling. In these applications, researchers will create a baseline forecast that attempts to predict how the world will evolve assuming current trends continue. This forecast is often called a "business-as-usual" case. Then, one or more alternative forecasts are created, to assess the potential effects of changes in key factors on the results. For example, an economic forecaster might use such an analysis to assess the likely effects of a change in property taxes on economic growth in a particular state.

THE WONDERS OF EXPONENTIAL GROWTH

One of the most basic concepts in forecasting is that of exponential growth. This concept generally applies to population forecasts because the rate of population growth is proportional to the number of people who are alive at any time (the more people there are, the more babies are born). Similarly, money in a bank account grows exponentially because the interest earned is related to how much money is in the account at any time.

Forecasts of population or economic activity are often summarized in the form of annual exponential growth rates. For example, the annual growth in gross domestic product was roughly 3.5% in 1996. If this growth continued at 3.5% per year for five years, total growth would be 1.035 x 1.035 x 1.035 x 1.035 x 1.035 - 1, or 19%.

Exponential growth can lead to surprising results in the longer term. **Table 30.1** compares three countries of 100 million people, one of which (Country A) grows two million people per year every year, and another of which (Country B) grows 2% every year (two million in the first year). After one hundred years, Country A will have grown to 300 million people, while Country B will have grown to more than 700 million! The differences become even more pronounced for higher growth rates. If Country C grew at 4%/year, its population would grow to more than five billion people a hundred years hence. A doubling of the exponential growth rate leads to a sevenfold increase in total population!

The famous futurist Julian Simon once made the statement that "We have the

TABLE 30.1. Exponential growth and population

	Units	*Country A*	*Country B*	*Country C*
Annual growth	Millions of people/yr or %/yr	2	2%	4%
Initial population	Millions	100	100	100
Population in 100 years	Millions	300	724	5050
Total growth over 100 years	Millions	200	624	4950
Average growth per year	Millions/year	2	6	50

technology to house, feed, and clothe a growing population for six billion years."[69] When told that any reasonable population growth rate would lead to the mass of all people exceeding the mass of the known universe over that period, he revised his statement to six *million* years. Even this revision wasn't enough to get Simon out of trouble, though. In six million years at modest growth rates, the number of people would exceed the number of electrons in the known universe!

Forecasts projecting exponential growth can lead to trouble when the rate of growth changes dramatically. During the 1950s and 1960s, growth in U.S. electricity use increased at a roughly constant 7%/year. Power planners had an easy job until the 1970s when the oil crisis, increasing energy efficiency, and problems with large power plants led to price increases and slowing demand for power. Growth rates in electricity demand slowed to 2–3% per year, and utilities were forced to cancel the planned construction of hundreds of power plants because of the drastically reduced growth in power use.

A KEY TRICK OF THE TRADE

In evaluating forecasts of exponential growth rates, an important trick is known as "the rule of 70." If you have an annual growth rate (say 2% per year for U.S. population) then the time it will take (in years) for population to double is roughly $\frac{70}{2}$ or thirty-five years. Growth of 7% per year implies a doubling time of about ten years. You don't need to know why this works, you just need to remember it.

Why is a doubling time useful? It is much easier to calculate multiples of two

than to actually calculate exponential growth in the formal way. When only an approximate number is needed, a doubling time is good enough. Say you want an approximate estimate of how much $10,000 now will be worth when you retire. If the growth rate after inflation is 7%/year (a reasonable long-term estimate for money invested in a tax-deferred stock account) the doubling time is ten years. In 30 years, the money will have undergone three doublings, or a factor of eight increase (2 x 2 x 2), and will be worth about $80,000 in 1997 dollars (of course, you'll have to pay taxes on that money when you withdraw it).

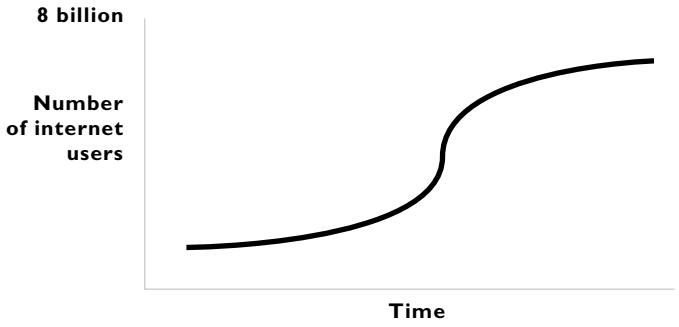
This approximation becomes less accurate as the growth rate increases. For rates above 10% per year, you'll probably want to check your results using more accurate methods, but for most situations, the rule of 70 is more than good enough (see *Ecoscience* for more details).

THE LOGISTIC CURVE

Many phenomena in nature and in business are described by the so-called "logistic curve" (also known as the "S curve"). This curve (shown in **Figure 30.1**) is characterized by slow but rising exponential growth in the early phase, rapid growth in the middle phase, and growth that tapers off as fundamental limits are reached (see *Ecoscience*). The logistic curve describes most phenomena eventually and applies equally well to technology adoption and population growth. We don't know where the top limits are in many cases, but in cases where we can identify those upper limits, the concept can come in handy.

One such example relates to the number of Internet users, which, at one point (in 1995 and 1996), was doubling every 100 days. At this rate, it would have reached four to eight billion people by 2002 or 2003, which is roughly the number of people in the entire world. This upper limit is a firm one that cannot be exceeded (barring the unlikely discovery of an alien world that wants to hook up to our computer networks). We can therefore infer that this growth will start leveling off over the next few years. The system may grow in other ways, of course, but, in this one aspect, the limits will not be exceeded. Knowing this fact can help a company that is planning its Internet business strategy cope with the future. The company can be sure that these upper limits will become important relatively soon, and, if its business model depends on continuous growth in numbers of users, the company will need to modify it.

FIGURE 30.1. A logistic curve



COPING WITH AN UNCERTAIN FUTURE

Forecasts are an integral part of the Cycle of Action. They are most useful in developing the action sequence, but they can also be important in the process of goal setting and in the evaluation of how actual events compare to our goals. In spite of their failings, and whether they are created using a formal model or developed in an intuitive way, they are an essential part of life and work.

3

28

If forecasts are part of your planning process, don't create just one. Instead, use a set of forecasts (i.e., scenarios) to explore the future. Vary key factors, and investigate which of them to ignore and which to dissect further. All forecasts are wrong in some respect, but if the process of designing them teaches you something about the world and how events may unfold, creating them will have been worth the effort.

26

Finally, it is important to adopt strategies that are robust in the face of inevitably imperfect forecasts. Several computer companies have moved to "build-to-order" manufacturing, which allows them to assemble computers as requested by customers. This strategy reduces dependence on forecasts but introduces other challenges in manufacturing (which are surmountable using current technology). This same lesson applies equally well to other such decisions: if the key variables are difficult or impossible to foresee, then use scenario analysis to evaluate the possible outcomes, and adopt strategies that are less dependent on forecasts.

26

EXERCISE

Find a forecast on a topic you care about, and learn how it was created. Talk to the author and read any supporting documentation. Do you still find it plausible?

In the space of 176 years, the Lower Mississippi has shortened itself 242 miles. That is an average of a trifle over one mile and a third per year. Therefore, any calm person who is not blind or idiotic can see that in the Old Oolitic Silurian Period, just a million years ago next November, the Lower Mississippi was upward of 1,300,000 miles long. By the same token, any person can see that 742 years from now, the Lower Mississippi will be only a mile and three-quarters long. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.

— **MARK TWAIN**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

31

HEAR ALL SIDES

One of the first rules of management in dealing with a conflict between coworkers is to listen carefully to all sides of the story before saying or doing *anything*. The same lesson applies in evaluating any two courses of action or points of view. It almost always pays to hear eloquent supporters of divergent viewpoints address each other's criticisms in a discussion or debate format. If you do not include such discussions as regular fodder for your deliberations, you are missing out on one of the most potent ways to clarify any set of issues. In rare cases, of course, such debates can muddy the waters still further, but they are almost always helpful as long as the participants are well intentioned, well matched, and prepared.

President John F. Kennedy knew this when he created a cadre of advisors who held widely divergent viewpoints during the Cuban Missile crisis. After the Bay of Pigs debacle, Kennedy vowed never to rely solely on his military advisors again. He created an ad hoc group of advisors, led by his brother Robert, to bring to his attention other alternatives. One of these options was the blockade of Cuba, which was ultimately the strategy that Kennedy adopted.

Such debates need not be formally structured, but structure can sometimes be an advantage. Formal debates in the Oxford style, sponsored by a British organization called "Intelligence Squared" and its US counterpart, have become popular recently (circa 2007). Each debate addresses a specific proposition like "global warming is not a crisis" with two skilled debate teams taking opposite sides of the issue. The audience is polled before and after the debate, and the team that manages to move more of the audience to its side than it had before the debate is declared the winner. Of course, the way that the proposition is initially framed can have a big influence on the quality of the discussion.

In the academic world, there is a tradition of "brown bag lunches" where researchers present work-in-progress to friends and close colleagues. These lunches help identify problems with analysis before a paper is written or pre-

sented to a wider group. They work best when all concerned are on friendly terms and the objective of clarifying the key argument is accepted by all participants.

The business world does not generally encourage the kind of interaction found in such "brown bags," to its detriment.[70] One acquaintance of mine works for a large mutual fund company. She lamented to me that the people under her supervision did not possess the ability or inclination to check their own work, which often led them to produce analysis that did not answer the management questions at hand. These analysts were responsible for creating reports that the firm used to assess financial fitness, so the reports were important. Unfortunately, her analysts had not learned what constituted good analysis, in part because the firm had no tradition of constructive intellectual criticism.

If you have reached a conclusion, it still pays to have a critic review your work. In fact, such review is virtually always a good idea (even at early stages of your analysis). Confirm your results using as many independent sources as you can find, just like you would test computer equipment before you put it in the field.

It is a capital mistake to theorize before you have all the evidence. It biases the judgment. — **SHERLOCK HOLMES**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

PART

V

SHOW YOUR STUFF

Don't forget the final step in any analysis: making your results useful to others. Your reports should be treasured by your colleagues as classics full of pertinent data, exhaustive documentation, and clear thinking. If they're not, you should follow the advice contained in these chapters and change your ways.

If you follow this advice, you'll soon gain a reputation for analytical integrity and clear thinking that can't be beat. If you don't, you'll be contributing to the piles of mediocre and barely used reports that inhabit bookshelves everywhere, and you most certainly don't want to do that.

They say that 90% of television is garbage. But 90% of EVERYTHING is garbage. — **GENE RODDENBERRY**

CHAPTER

32

KNOW YOUR AUDIENCE

How results are expressed is at least as important as what the results are. Most analysts are so involved in their calculations that they forget that other people don't care nearly as much about the results of the analysis as they themselves do. Good analysis is often handicapped in its presentation because the authors have not taken the vital last step of thinking about what the audience cares about and expressing the results in those terms.

One of my favorite professors from graduate school, Art Rosenfeld, is a master at making complex analysis results understandable to the lay person. Art presented his estimates of energy savings from the latest efficient gadget (a new refrigerator, for example) for many different people. Sometimes he would testify before congress, other times he would address public utility commissioners, and still other times he would tell his story for a general audience. He would always tailor his talk to the listeners. If oil drilling in the Arctic National Wildlife Refuge was a hot topic before congress that day, he would express his energy savings estimates in terms of "Arctic Refuges saved." If the public utility commissioners were concerned about canceling a power plant construction project, Art would calculate how much the savings were worth in terms of "avoided power plants." And for the lay audience, Art would show how much money people would save per household if they bought the better fridge. In each case, Art's message gained resonance because he thought seriously about the concerns of the audience.

In both talks and papers, your audience will only retain a few key points from your work. *You* must decide what those few points should be, and hammer them home. If you don't, you risk not reaching the people you set out to convince. So step back and *reflect* after you've completed your analysis. Think about what you've learned, and explain it in three or four simple sentences. *Dig into the numbers* to support your main points, but don't overwhelm your audience with numbers. Instead, focus on the conclusions, describe why your

21 27

analysis makes such a compelling case, and explain it in ways specifically tailored to reach your audience.

This lesson is particularly important when the analysis results are intended to affect public opinion or a public policy debate. Journalists generally latch on to only a few key results of a complex study when deciding what "angle" to take with a story. Express these results to reach out to the desired audience, and your effectiveness will increase many times over.

In order to achieve victory, you must place yourself in your opponent's skin. If you don't understand yourself, you will lose one hundred percent of the time. If you understand yourself you will win fifty percent of the time. If you understand yourself and your opponent you will win one hundred percent of the time. — **TSUTOMU OSHIMA**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

33

DOCUMENT, DOCUMENT, DOCUMENT

Over the years, friends have clipped dozens of articles for me on topics in which they knew I had an interest. In almost all cases, they neglected to write on the clipping the full reference, thus rendering it unusable for research purposes. An article without a reference is useless except as background info. An article with a reference is a citable source.

Documentation lends *credibility* to any intellectual effort. It also recognizes the previous work of others and allows readers to follow your thought processes (it also allows YOU to recreate your thinking months after you have achieved some conceptual breakthrough). Any competent analyst ought to be able to recreate your analysis from the documentation you provide, and you must be able to do the same more quickly than others can. Finally, the process of documenting your results can help you check those results and ensure accuracy.

GIVING CREDIT WHERE CREDIT IS DUE[71]

Progress in intellectual endeavors depends on prior work being fairly acknowledged. When you have relied on a particular fact or insight that you got from another source, you must reference it. You can put this reference in a footnote, you can use the (Author Date) method (e.g., (Koomey 1995)), or you can consult *The Chicago Manual of Style* for more reference formats.

If you quote from another source, cite the source (and go back to the original source if you can). If it is an extensive quotation, if the quotation represents the "heart" of the work, or if using the quotation might affect the commercial value of the cited work, it may be appropriate to obtain the author's permission for using it. "Fair use" has a specific legal definition, but there are many gray areas. You are allowed to quote excerpts from other people's copyrighted work to illustrate, clarify, or comment on their work. Whether an excerpt constitutes

fair use depends on the context in which the excerpt is used and on a variety of other factors. One common-sense test to apply is to ask yourself "If the cited source were my work, how would I feel if someone else cited it in the way I cite it here?"

For more details on copyrights in the digital age, see the Copyright Office web site for the Library of Congress <http://www.copyright.gov>, and the Nolo.com site devoted to this topic <http://www.nolo.com/>. For information on obtaining permission for the use of copyright protected material, see Richard Stim's excellent book titled *Getting Permission*, also available from Nolo.com.

The Internet has made issues of credit and authorship more complex. One friend of mine occasionally teaches a class on environmental science, and he found that in term papers submitted by his students, three-quarters of all the sources cited were World Wide Web addresses (called Uniform Resource Locators, or URLs), with little or no explanation given for how the information was derived or whether it was credible. For the students, the mere existence of a URL was enough reason to cite it when in fact more investigation and documentation is really required to ensure that the web site gives proper credit and accurately reports results.

If you do rely on a web site for a research project, don't forget to make an electronic file (PDF) or a hard copy printout of the site. Web sites change all the time, and their impermanence places special responsibility on the researcher to record the information in a permanent way. Another important detail is to set your web browser to include the actual URL in the header or footer of any printed copies of web sites. Some browsers do this automatically, but yours may not, so make sure this setting is saved as the default option.

LEAVING A TRAIL FOR YOU TO REMEMBER

The fancy technical name for data documentation is "source tagging." Researchers who specialize in data quality focus on how to decide which information about a source is appropriate for electronic data and how to attach this key information to those data. I refer to data without attached sources as "disembodied." I also call them useless, because without the sources, they are little better than rumors.

Don't underestimate the importance of good archiving practice: Archiving is the process by which older data and records are stored for retrieval if needed. Companies neglect this process at their peril because mission-critical information must be accessible to those making strategic decisions. This task takes on greater importance in the electronic age because the "paper trail" that existed previously is now commonly absent (see Cook's 1995 *Technology Review* article for details).

Keep a notebook for each major project: Record in this book important insights you have, notes from meetings, reports to track down, or any other items related to the progress of your work. An electronic organizer serves this function, but if you haven't adopted one of these handheld devices, then a paper notebook will serve just fine. These notes may come in handy later if you want to build on previous work.

Always date your work, even your roughest handwritten notes: Reconstructing the evolution of your thoughts often requires knowing *when* you wrote something down.

Always write down the full reference to an article that you've copied, as soon as you've copied it: Write it on the first page of the copied article. No exceptions! Until I learned this lesson, I can't tell you how many times I had to go back and find the source because I had not written down the reference at the moment I copied the article.

Always write down why you think a particular article is significant for your work, because in six months you probably will have forgotten: Any longer-term research project involves more details than you can keep in your head, so keep good notes on articles you decide to clip.

Be selective about the articles you clip and file, but clip the important ones without fail: In this age of ubiquitous web access, electronic databases of most publications are widely available. Save time and avoid clipping articles unless they are immediately useful for a current research project. Do print and file those articles you actually use, so you can refer to them later.

7
9

If you create computer code or spreadsheets, put extensive explanations of your logic into the code in comments and footnotes: Describe your reasoning as if someone else will be reading that code and trying to understand it because that is literally the case: in six months, you will be a very different person who has forgotten the details that led you to structure your calculations in the way you did.

If you work with spreadsheets, never type data directly into spreadsheet formulas: For example, some people enter formulas that look like this into a spreadsheet cell: = 2 x 457.3/9415 + 24. Instead, enter these numbers into cells, then create formulas that use these cells for calculations, like this: = F2 x F3/F4 + F5 (in this case, the various numbers are in cells F2 through F5). You can easily change the data later without having to fiddle with the formulas, plus you can see the input data for the calculation without having to poke around in the cell containing the formula. Such planning ahead can save you time and make it easier to spot mistakes in the data and calculations.

Segregate input data in one part of the spreadsheet and results in another part: Be systematic and you will be able to recreate your calculations even a year from now, after you've forgotten the assumptions that are now so fresh in your mind.

Define units: Make sure your documentation is clear about exactly which units you are using. For example, the word "ton" has at least five different possible meanings. There are short tons (2,000 pounds or 907 kg), long or deadweight tons (2,240 pounds or 1,016 kg), metric tons (sensibly defined as 1,000 kg), tons of cooling (defined as 12,000 Btus/hour of cooling capacity) and register tons, which are (oddly enough) defined in units of volume! Also beware of definitions that vary by country. The word "billion" means 1,000,000,000 (10^9) in the U.S., but in much of Europe that same number is written as "thousand million." In Europe, a billion is 1,000,000,000,000 or 10^{12} (which in the U.S. is called a "trillion"). For the sometimes confusing history of the word "billion," see the *Wall Street Journal*'s "Numbers Guy" column of March 30, 2006. For an excellent on-line dictionary of units, see <http://www.unc.edu/~rowlett/units>.

Define constants: Constants and conventions must be carefully defined, or confusion can result. For example, analysts typically assume that a year is equal to

8,760 hours, which is 365 days times twenty-four hours per day. In reality, the average year is 8,766 hours, because of leap years (every fourth year has an additional day). The small difference in these two numbers may lead to slight discrepancies, but other such conventions may have larger effects.

Go metric: Use metric units in any analysis reports that may be useful to readers outside the U.S. If your U.S. audience would find those units confusing, include *both* English and metric units. Because the World Wide Web is an international medium, you should ensure that your work will gain wide use by those in other countries who use different (and in this case more logical) conventions than you do. Tailoring your results for the ease of these readers will help your message spread more quickly than it would otherwise.

38

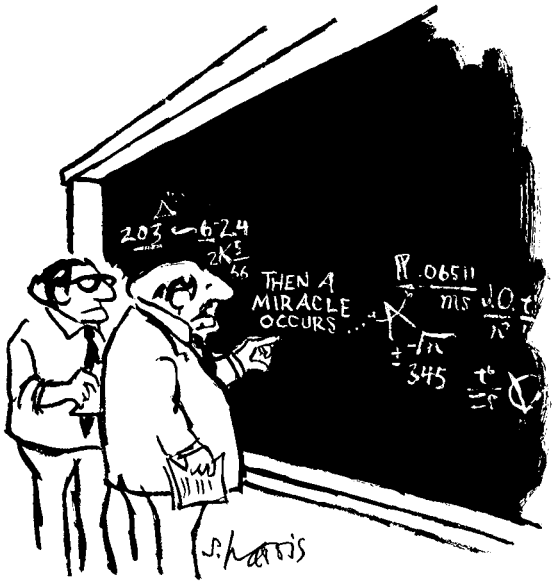
32

CREATING A TRAIL FOR OTHERS TO FOLLOW

The importance of documenting your work comes to the fore when there is some kind of a crisis. In August 1997 my former boss, Mark Levine, was called before the President's Council of Economic Advisors to testify about a report we had recently completed. I had created most of the calculations summarizing the results and had written a detailed description of how the calculations were done. I crafted the documentation for a technical analyst who wanted to reproduce my calculations and who had time to delve into depth. While Mark is a technical analyst, it soon became obvious that he needed something a little different because his time was so limited. I talked him through the calculations, and the testimony went well. The lesson was not lost on me, however.

Assume that the person reading your documentation has limited time to figure it out. This is the right approach when documenting because it will ensure that you take appropriate care. You may think that your task won't require such thoroughness, but what if someone else has to pick up where you left off?

Describe the calculations in such excruciating detail that, even in the worst case, an intelligent reader could understand and reproduce your calculations as quickly as possible. Put a clear description of your methods in the main body of your report; then, relegate the complete documentation to appendices. Such detail may seem like overkill at the time, but you will never regret it. You may not be able to anticipate how your calculations will be used in the future.



"I think you should be more explicit here in step two."

© 2007 Sidney Harris from cartoonbank.com. All Rights Reserved.

10 Documentation is one of Stephen Covey's "important but not urgent" tasks and one that should be a priority in any analysis work.

Your references to other people's work must be complete so that others can follow up. References must contain enough information to enable a reader to find a reference at the library even in the face of the inevitable typos and smudged copies. Redundancy is good in this case. **Table 33.1** shows the information you should include for the most commonly referenced sources. How you choose to present this information in your bibliography is up to you, but I recommend following the basic formats in the *Chicago Manual of Style* (a nice summary of this and other reference formats is contained in *A Writer's Reference*, by Diana Hacker).

Computer programs that automate the process of tracking references (such as EndNote) are one of the few innovations since the word processor to increase significantly the productivity of writers and researchers. Use them if you do more than a little writing. These tools can ensure that your references are complete and do so in a fraction of the time and annoyance that in the past accompanied compiling a bibliography. These tools allow you to create one database

TABLE 33.1. Information needed for complete references

Type	*Author*	*Year*	*Title*	*Institution*	*Volume*	*Number*	*Page #*	*Date*	*URL*	*Other*
Book	Yes	Yes	Yes	Publisher	Edition		If applicable		Yes	City
Edited Book	Yes	Yes	Yes	Publisher	Edition		If applicable		Yes	City
Book Section	Yes	Yes	Yes	Publisher	Edition		Yes		Yes	Book title, City
Journal Article	Yes	Yes	Yes	Journal name	Yes	Yes	Yes	Yes	Yes	
Magazine Article	Yes	Yes	Yes	Magazine name	Yes	Yes	Yes	Yes	Yes	
Newspaper Article	Yes	Yes	Yes	Newspaper name			Yes	Yes	Yes	Section, City
Report	Yes	Yes	Yes	Name of institution			If applicable	Yes	Yes	City
Conference proceedings	Yes	Yes	Yes	Conference organizer	Yes		Yes	Yes	Yes	Location
Thesis	Yes	Yes	Yes	University			If applicable	Yes	Yes	Thesis type, University location
Personal Communication	Yes	Yes	Yes	Name of institution			If applicable	Yes	Yes	Address, phone #, fax #, email

of your references, and, once you've entered them into the database, you need never enter them again. It formats the references in any way you choose and comes with a collection of formats for many of the more common academic journals. Acquire reference database software with all due haste!

CHECKING YOUR WORK

You wouldn't put a computer system into a client's office without first testing it, would you? Why then do so many people produce analysis without checking the results for internal consistency, completeness, and congruence with the stated goals of the people requesting the analysis? An important function of the documentation process is to help conduct this testing. Explaining your methods in detail helps identify conceptual blunders. If you can't explain the reasoning behind your methods, then you have more work to do.

Documentation should be an integral part of your analysis process. As you are

22 conducting your research, you should write down your methods or try to explain them to a colleague. Your analyses will be the better for this give and take.

IF YOU ARE WORKING WITH DOLLARS

The value of money is affected by inflation. Over time, inflation makes each dollar worth less, so a dollar spent in 1997 is *not* the same as a dollar spent in 1990. You must, therefore, document the year in which expenditures occur.

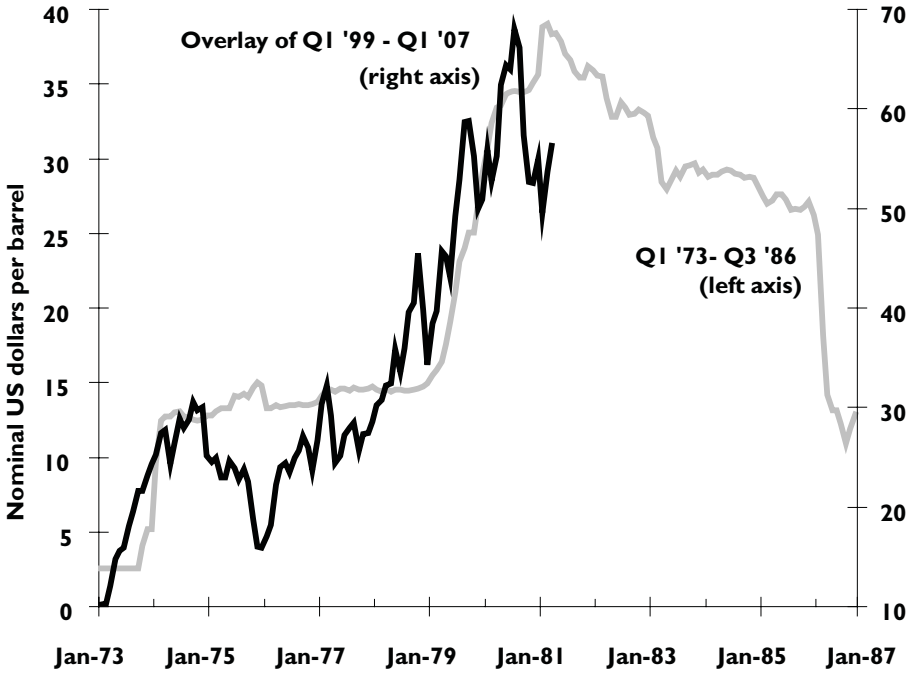
Any time you present tables or figures that include dollars, you *must* specify "current dollars" or "constant dollars in some year." Current (also known as nominal) dollars imply that if the money is spent in 1990 it is in 1990 dollars and if spent in 1997 it is in 1997 dollars. Alternatively (and more accurately in my view), you can correct for inflation using an inflation index from the **9** *Statistical Abstract of the U.S.* to calculate all expenditures in constant dollars. The most common indices for this purpose are the Consumer Price Index and the implicit Gross Domestic Product (GDP) deflator (sometimes called the "chain-type price index for GDP"), but sometimes a more specialized index for a particular material or industrial sector comes in handy. The expenditures in current dollars in 1990 would then be converted to 1997 dollars by multiplying them by the following factor:

$$\frac{\text{Inflation index in 1997}}{\text{Inflation index in 1990}}$$

If the inflation index in 1997 is 1.5, and the inflation index in 1990 is 1.2, then expenditures in 1990 must be multiplied by $\frac{1.5}{1.2}$ or 1.25 to convert them to 1997 dollars. All other years would be treated similarly using the appropriate inflation index (of course, 1997 expenditures are already in 1997 dollars).

Correcting for inflation, although simple, is commonly overlooked. When the U.S. Postal Service announced in May 1998 that it would be granted an increase in the price of a first-class stamp in 1999, *The San Francisco Chronicle* presented a summary of the new rates and printed a graph of first-class postal rates since 1885.[72] This graph showed a huge increase in prices, from 2¢ in 1885 to 33¢ in 1999. An uninformed observer might easily have concluded from this graph that the price of first-class mail had skyrocketed, particularly since the 1970s. The problem, of course, is that the graph was not corrected for inflation,

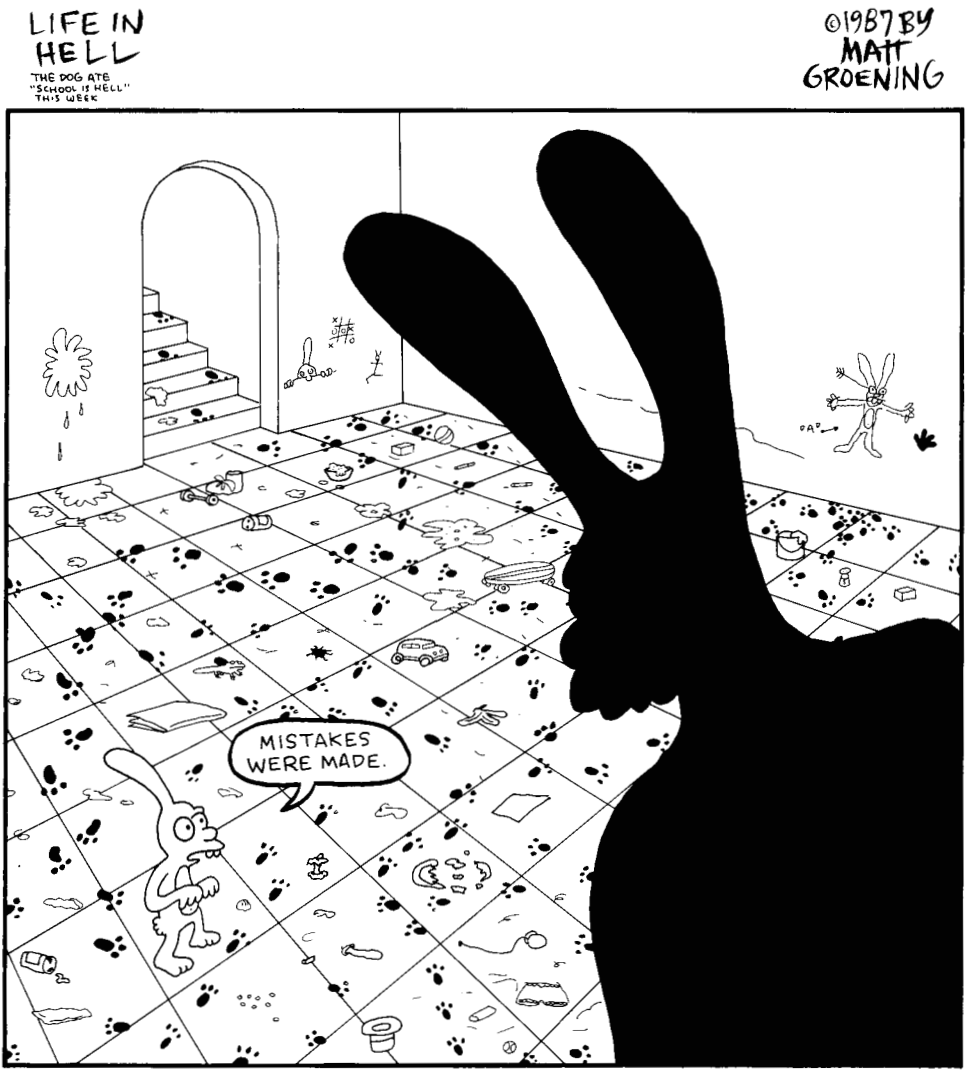
FIGURE 33.1. Nominal refinery acquisition cost for imported crude oil to the US



NOTE: Price data taken from the US Energy Information Administration.

and nowhere did the *Chronicle* article mention this omission. *The Oakland Tribune* story on the same topic[73] correctly reproduced two graphs, one in current dollars and one in inflation-adjusted dollars. The second graph showed that the 33¢ price of first class mail in 1999 actually represented a decline in constant dollars from 35¢ in 1885.

A more recent example appeared in the investor newspaper *Barron's* in September 2007[74] (the *Barron's* graph plotted the spot price of Saudi light crude oil, but the picture is about the same using the acquisition price for crude oil imported to US refineries, which is what I've plotted in **Figure 33.1**). I will have a lot more to say about the incorrect and misleading attributes of this graph in the chapter on improving graph design, but for now, just note that the reader was asked to compare two segments of the time series of oil prices and marvel at how similar they looked. "I've seen this movie before," said the person being interviewed about his predictions about future oil prices, implying that the similarity in the two graphs made a subsequent drop in oil prices likely. Regardless of the merits of that argument (and there are some) it is in this case based on false prem-



From *The Big Book of Hell* © 1990 by Matt Groening. All Rights Reserved. Reprinted by permission of Pantheon Books, a division of Random House, Inc., NY. Courtesy of Acme Features Syndicate.

ises: It is simply incorrect to compare nominal prices from two different time periods without correcting for inflation. Even a normally excellent business publication like *Barron's* can fall into this trap, but you need not be misled.

I once found a report that contained information on the costs of computing power in the 1950s and 1960s, and at first I was delighted. My delight turned to dismay when I realized that the report did not document whether the costs were current or constant dollars, and I was unable to use the information because I couldn't determine how to treat the costs. What a waste! Don't let this happen to your work.

There are similar (but even more intricate) subtleties regarding the use of currency exchange rates to convert foreign currencies to dollars or vice versa. Because such rates fluctuate over time, you must document which exchange rates you are using; otherwise your calculations are useless to the reader.[75]

TAKE RESPONSIBILITY FOR YOUR WORK

When documenting or writing reports, take responsibility by using the active voice. For example, say "I estimated the following parameters" instead of "the following parameters were estimated." Make it crystal clear who did what, so there is no ambiguity. The passive voice is the tool of people who want to avoid responsibility for their actions, but the purpose of documentation is to assign responsibility, for all the reasons listed above. Don't thwart your efforts by using the often ambiguous passive voice.

DOCUMENTATION PAYS OFF

Any time there is a new "hot" analysis topic, people with little prior knowledge jump in. You can perform an immense service to any debate (and garner attention for your point of view) simply by being systematic in identifying and investigating such topics and documenting them in a careful way.

A case in point is the raucous debate over the cost of pollution to society from power plants during the late 1980s and early 1990s. In 1990, I wrote a paper that was not a conceptual breakthrough but simply created consistent comparisons among different estimates of these social costs. I received hundreds of requests from all over the world for this admittedly rather dry technical report as well as offers to speak at conferences and opportunities to contribute to other reports on the topic. It garnered attention in large part because of its systematic and careful documentation. You too can use this approach to your advantage.

The same sort of intellectual chaos that pervaded that debate now characterizes the emerging demand for advertising and commerce on the web. Businesses want hard numbers about the effectiveness of web advertising in selling their products, but solid statistics are still scarce and controversy reigns about different methodologies. Systematic thinking in this area could be of

immense value to the companies deciding on advertising budgets. As of this writing, the decisions are being driven more by hype than data.

BACK UP YOUR DATA!

Many analysts don't routinely back up their electronic files. This oversight is potentially disastrous, as the otherwise technology-savvy Stanford Business School found out in April 1998. Two large computers were moved before the operators verified that all data had been backed up, and almost a dozen researchers lost years of work.[76] Hard disks, floppy disks, and data tapes don't last forever, and they are guaranteed to fail at the least opportune time (they can also become obsolete as technology changes). I have many colleagues who have lost data from such failures, and it's never fun. So after you've gone to the trouble of carefully documenting your work, make sure that there are at least a couple of electronic copies, and put a paper version in a safe place. One trick is to email electronic files to yourself, which places them on the server of your Internet service provider and adds an additional layer of redundancy. Also don't forget that data formats change over time, so if you have some older files, convert them to a new format before trashing the program that created them.

CONCLUSIONS

Ignorance of good data documentation practices is widespread. These skills should be taught in high school, yet technically trained college graduates routinely do not identify their data sources or the methods they used for their calculations. Don't fall into this trap. Make sure that any intelligent person (including YOU!) can reproduce your analysis from the footnotes. With this one simple step, you will vault above 95% of the world's analysts. Leaping over the other 5% depends on the quality of your content.

Results! Why, man, I have gotten a lot of results. I know several thousand things that won't work. — **THOMAS A. EDISON**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

34

LET THE TABLES AND GRAPHS DO THE WORK

Don't just write to be understood; write so that you cannot be misunderstood. — **ROBERT LOUIS STEVENSON**

In any writing that relies on quantitative analysis, the calculations are the bulk of the work. First, create the tables documenting data and presenting results; then, write the text describing the inputs and analysis. For particularly important results, make graphs that best emphasize the points you want to make. **36** **35**

The text should call out only the most important numbers in the table for ease of comprehension. The numbers junkies can delve into the table for any details they feel compelled to unearth.

The following passage is an example of the right way to summarize technical analysis, taken from the section "Dig into the numbers" above: **27**

> **Table 34.1** summarizes the key financial parameters for calculating market capitalization for the two companies.
>
> The first thing to notice about the financial statistics for these two companies is a huge disparity. While GE's market capitalization in November 2007 was about twice as large as Google's, GE's revenues were almost twelve times larger. All other things being equal, we might expect that companies with similar market valuations would also have similar revenue streams. All other things are not equal, however, and finding out why will help illustrate this important analytical technique.
>
> If Google's market capitalization per dollar of revenues was the same as GE, we would expect that its market capitalization would be only one-sixth as large as it is and we need to explain this six-fold discrepancy. To do so, we examine the components of the last equation above. The first component

TABLE 34.1. Comparison of financial statistics for Google and General Electric (GE)

	Units	*GE*	*Google*	*Ratio Google /GE*
Annual revenues	B$ 2007	174	15.0	0.086
Number of shares	Billions	10	0.3	0.03
Price per share	$2007/share	40	722	18
Diluted Earnings per share (EPS)	$2007/share	2.2	12.8	5.9
Trailing price-to-earnings (PE) ratio	Ratio	19	56	3.0
Total earnings	B$ 2007	22	4.0	0.18
Earnings per dollar of revenues	$/dollar	$0.13	$0.26	2.1
Market capitalization	B$ 2007	412	226	0.55
Market cap. per dollar of revenues	$/dollar	$2.4	$15	6.4

(1) B signifies billions (thousand millions for British readers).
(2) PE is the share price to earnings ratio (measured as an intraday value).
(3) Market capitalization (measured as an intraday value) is the product of # of shares and price per share. Price per share is the product of EPS and PE ratio.
(4) Values taken from Yahoo! Finance on November 4, 2007.

is earnings per dollar of revenues, and the second component is the price-to-earnings ratio.

As Table 27.1 shows, Google's earnings as of November 2007 were about twice as big as GE's per dollar of revenues, which accounts for about a factor of two in our six-fold difference. The price-to-earnings ratio also differed between the two companies. Apparently, the stock market valued one dollar of Google's earnings three times as much as one dollar of GE's earnings, which accounts for the remaining difference.

Now compare it to the text below that doesn't follow the approach I advocate:

Table 34.1 summarizes the key financial parameters for calculating market capitalization for the two companies.

The first thing to notice about the financial statistics for these two companies is a huge disparity. While GE's market capitalization in November 2007 ($412B) was about twice as large as Google's ($226B), GE's revenues were almost twelve times larger ($174B vs. $15B). All other things being equal, we might expect that companies with similar market valuations would also have similar revenue streams. All other things are not equal, however, and finding out why will help illustrate this important analytical technique.

> If Google's market capitalization per dollar of revenues ($15) was the same as GE's ($2.4), we would expect that its market capitalization would be only one-sixth as large as it is, and we need to explain this six-fold discrepancy. To do so, we examine the components of the last equation above. The first component is earnings per dollar of revenues, and the second component is the price-to-earnings ratio.
>
> As Table 27.1 shows, Google has 0.3 billion shares outstanding, with earnings of $12.8 per share. GE has about 10 billion shares in the hands of stockholders, with earnings of $2.2 per share. The product of the number of shares and the earnings per share gives the total annual profits for each company. For Google, profits are about $4B on revenues of $15B, and for GE, profits are about $22B on revenues of $174B.
>
> Every dollar of Google's revenues therefore yields twice as much profit as does a dollar of GE's revenues ($0.26 compared to $0.13 per dollar of revenues). So the increased profitability of Google per dollar of revenues accounts for about a factor of two in our six-fold difference.
>
> The price-to-earnings ratio also differed between the two companies. Apparently, the stock market valued one dollar of Google's earnings (PE = 56) three times as much as one dollar of GE's earnings (PE = 19), which accounts for the remaining difference.

This text is potentially intimidating to the reader who is less comfortable with numbers, it is less compact, and it repeats data that are already presented in the table itself. It also requires much more exacting proofreading

In addition, if the calculations change, you'll have to manually change the text to reflect those calculations. If the calculations are in a table printed or copied from a spreadsheet, you merely change the calculations in the table, print it out again or copy it into your text, and change the few key numbers in the text that refer to the table. This approach is far less labor intensive when you need to make changes (and you almost always need to make changes at some point).

If you omit the table and just include the more numbers-heavy text, as is commonly done, you make it even harder for readers to assess the validity of your assertions. The table organizes the information and makes it conveniently available to the reader. It also makes it easier to write the text.

The main purpose of writing of this type is to present, in as clear a fashion as possible, the logic of the calculations, the main input data, and the key

results. Any competent analyst ought to be able to reproduce your analysis
33 using the documentation contained in your report. If they can't, your work isn't done. An astounding number of analysts routinely omit vital data and assumptions from their reports, a practice for which there can be no excuse.

If the analysis is well conceived, the tables and graphs well designed, and the audience clearly defined, the report should practically write itself. I normally use the following main headings in technical reports (your specific needs may dictate a slightly different structure):

Introduction
Methodology and Data
Results
Discussion
Limitations of this analysis
Future work
Conclusions
References
Appendices

Using this method, the writing task becomes one of straightforward description, except in the "Discussion" and "Conclusions" sections, where you describe why the results are important.

Place detailed calculations and supporting data in appendices. Only put *essential* information in the main body of the report. When possible, write relatively short technical reports that summarize the analysis results and relegate complete documentation to appendices. It is then a simple matter to take the short report, eliminate all but the three to five of the most crucial graphs and tables, change some text, and send it off to a journal for publication.

One trick that helps the reader of longer articles and reports is to put in bold face the first mention of tables and graphs in the text so it is obvious where they are discussed and explained. For example: "**Table 1** shows the results . . ." This trick makes it easier for readers who jump straight to the results and want to comprehend the analysis quickly. There are many such readers, which also
33 makes it essential that each table and figure have footnotes that document the key elements of the supporting calculations.

If the reader needs to know a precise value, then use a table. If the shape of the data conveys insights, then make a graph. Use graphs to tell stories and use tables to supply data. Choose the means to communicate your messages most effectively.

There may occasionally be good reasons for breaking the rules described above. However, by following them most of the time, and by thinking carefully about the concerns of your intended audience, you will most effectively convey the implications of your results.

32

The design of the data structures is the central decision in the creation of a [computer] program. Once the data structures are laid out, the algorithms tend to fall into place, and the coding is comparatively easy.

— **BRIAN W. KERNIGHAN AND ROB PIKE**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

35

CREATE COMPELLING GRAPHS AND FIGURES

Modern graphs are largely an eighteenth-century creation. Edward Tufte, one of the modern masters of graphical display, states that Playfair, Lambert, and others developed the first graphs that broke away from a geographical representation. In the two centuries since, such charts have become commonplace. Modern computer tools make creating graphs from data simple—in fact, it has probably become too easy. Graphs are often produced without thought for their main purpose: to enlighten and inform the reader.

Tufte defines graphical excellence as "that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space."[77] He promotes several rules that will help you avoid the most common pitfalls, which I summarize below:

- "*Above all else show the data":* Make sure the graphs highlight your results instead of hiding them as many graphs do. Use the graphs to tell stories about your analysis, and make those stories leap out at the reader.
- "*Maximize the data-to-ink ratio, within reason":* Tufte defines the data-ink ratio as the amount of ink that conveys information divided by the total amount of ink used for the graph. Many graphs have low data-ink ratios, and you should avoid this pitfall. To maximize the data-ink ratio, "erase non-data-ink and redundant data-ink."
- "*Eliminate chart junk":* Remove clashing fill patterns (Moiré patterns), which are hard on the eyes. Eliminate extraneous gridlines, which obscure the data. Delete graphical elements that decorate but convey no information, such as 3-dimensional (3-D) chart elements on 2-D graphs.

  This latter mistake is a common one, particularly in "slick" brochures and company annual reports. Although there are legitimate uses for 3-D graphs, a bar chart is not one of them. I adapted **Figure 35.1** from the 1999 annual report of a company with $6 billion in sales in 1999.

18 32

FIGURE 35.1. The wrong way to make a bar chart

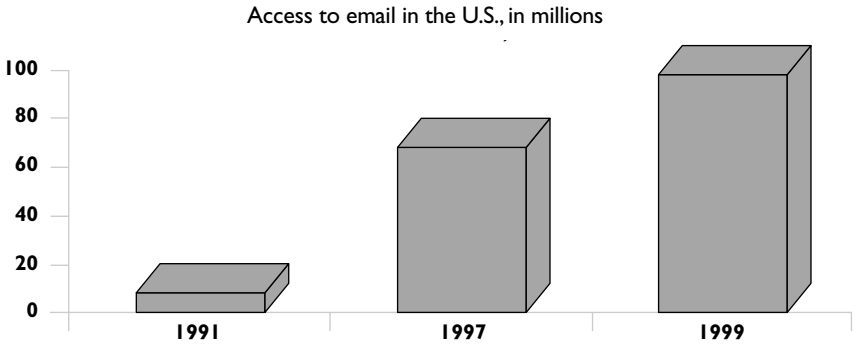
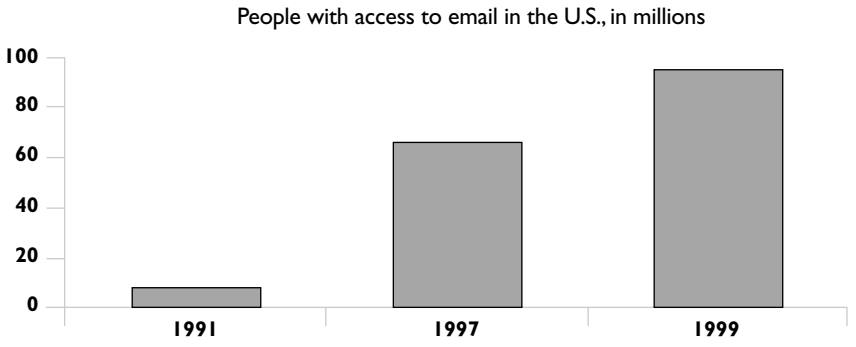


Figure 35.1 shows access to email in the U.S. using 3-D bars. Because of the 3-D perspective, it is impossible to extract the data accurately from the graph. **Figure 35.2** shows how the graph should look for greatest accuracy.

FIGURE 35.2. A better way to make a bar chart



- "*Maximize data density" (numbers per unit area), within reason:* Typical graphs in newspapers and journals, based on Tufte's review of hundreds of charts published 1970-1980, have data densities of 5 to 20 numbers per square inch (0.8 to 3 numbers per square centimeter). The densest graph Tufte examined (a map of galaxies contributed by some astronomers) had a data density of 110,000 numbers per square inch (17,000 numbers per square centimeter); the least dense graph he considered had a data density of 0.15 numbers per square inch (0.02 numbers per square centimeter). Maps tend to have high data density; badly designed charts tend to have the lowest.

  When things get really bad, a graph may contain only one data point.

I encountered such an appalling graph on an airplane flight from San Francisco to Washington, DC on June 8, 1999. It showed the size of the market for an electronic gadget, and it looked like **Figure 35.3**:

FIGURE 35.3. The market for an electronic gadget

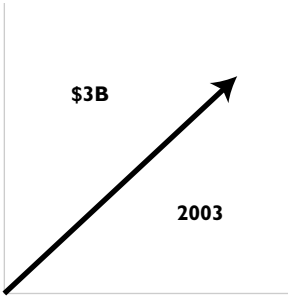
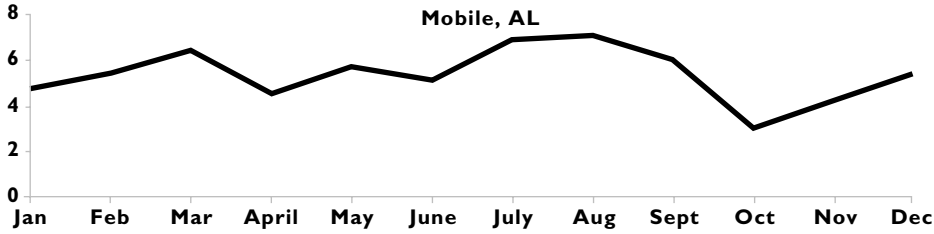


Figure 35.3 was all that appeared on a 17" diagonal TV screen, which represents a data density of 0.008 numbers per square inch (0.001 numbers per square centimeter), probably an all-time low. It gives no indication of the current size of the market except that the arrow going up and to the right implies that the market is growing from today until 2003. Avoid this blunder and maximize data density.

- "*Use small multiples":* One important way to maximize data density is to create what Tufte calls small multiples, defined as "a series of graphics, showing the same combination of variables, indexed by changes in another variable."[78] Let's say you plot monthly rainfall for Mobile, Alabama, as shown in **Figure 35.4**. If you have data on rainfall for many cities that you'd like to compare, you can shrink them and plot them side by side, as shown in **Figure 35.5**. You need not reproduce all of the labels on every chart. Once the reader has learned the "key," small multiples like these allow quick and easy comparison of large amounts of data. For many more examples of this technique, see Chapter 11 of *Show Me the Numbers* ("Design Solutions for Multiple Variables").
- "*Revise and edit":* All work improves with revisions. Leave enough time to get feedback and make your graphs better.

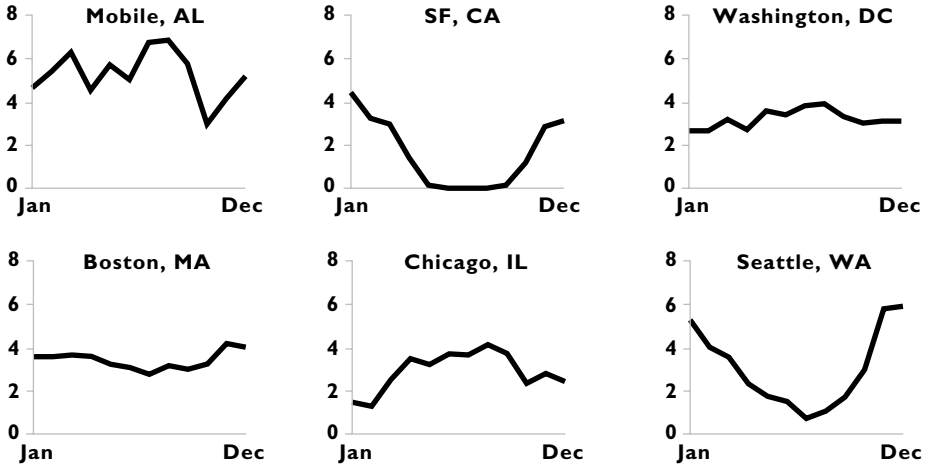
I add one more rule of my own, which is *Make figures stand alone*. Figures are often copied separately from the report in which they appear, to allow the reader to trace your figure back to the source. Put the date and file name on the figure, at least when it's in draft form. Make sure the final version contains some reference to the report in which it appears or to the author's contact informa-

FIGURE 35.4. Average monthly rainfall for Mobile, Alabama, 1961-1990 (inches)



Source: *Statistical Abstract of the U.S. 1997*, Table 397.
Normal Monthly and Annual Precipitation–Selected Cities (based on 1961-1990 data).

FIGURE 35.5. Average monthly rainfall for selected cities, 1961-1990 (inches)

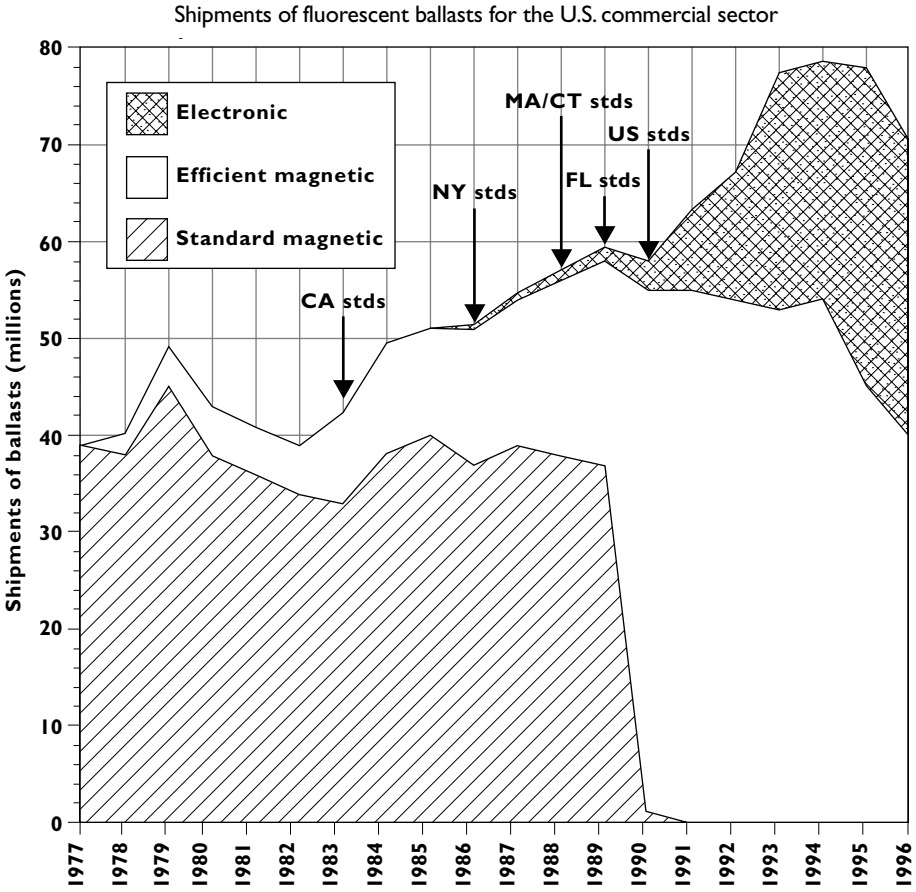


tion. Figures are often copied in black and white even if the original graph is in color. Make sure a black and white copy conveys the same message as the original, and give the reader the report and author's contact information on the graph as a backup in case the colors don't come through in the reader's copy.

APPLYING THESE RULES

Let's look at a figure I created in the past and make it better using these rules. **Figure 35.6** shows a graph of the historical annual sales of fluorescent lamp ballasts for U.S. commercial buildings that appeared in a 1997 report.[79]

FIGURE 35.6. A graph that needs work

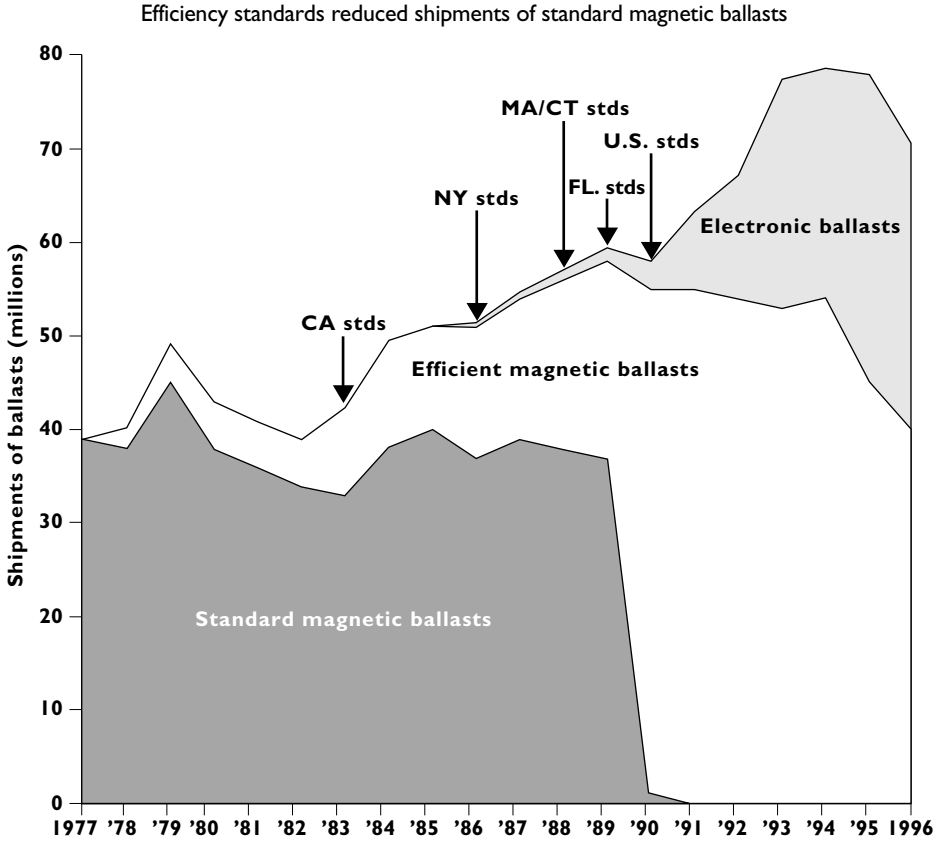


NOTE: Because they are not typically used in the commercial sector, the following ballast types are omitted from the data presented in this figure: low-power-factor ballasts (shipments of 24.2 million in 1996), 1500 MA high-power-factor ballasts (0.2 million in 1996), and ballasts included in the "All other corrected power factor" category of the Census Bureau data. Ballast imports are not included because they were not disaggregated by ballast type in the census data. Arrows indicate the years in which state standards and, finally, federal standards took effect, preventing the sale of standard magnetic ballasts.

SOURCES: Koomey et al. (1995); Census Bureau MQ36C(94)-5, Table 3; Census Bureau MQ36C(96)-5, Tables 3 and 4

Notice that the patterns used in the area chart seem to "vibrate" when you stare at them. This is what Tufte calls the Moiré effect, and it's best avoided. The gridlines are extraneous, and the legend can easily be eliminated. The years along the X-axis are a bit hard to read because they are turned 90 degrees from horizontal. The footnote is descriptive and complete, but the figure still lacks the contact information for the author of the report and the file name of the graph.

FIGURE 35.7. An improved graph



NOTE: Because they are not typically used in the commercial sector, the following ballast types are omitted from the data presented in this figure: low-power-factor ballasts (shipments of 24.2 million in 1996), 1,500-MA high-power-factor ballasts (0.2 million in 1996), and ballasts included in the "All other corrected power factor" category of the Census Bureau data. Ballast imports are not included because they were not disaggregated by ballast type in the census data. Arrows indicate the years in which state standards and, finally, federal standards took effect, preventing the sale of standard magnetic ballasts.
SOURCES: Koomey et al. (1995); Census Bureau MQ36C(94)-5, Table 3; Census Bureau MQ36C(96)-5, Tables 3 and 4.
CONTACT: Jonathan Koomey, JGKoomey@lbl.gov. <http://enduse.lbl.gov>
FILENAME: Final Ballast Charts
DATE OF CREATION: February 1995.
DATE OF LAST MODIFICATION: October 1997.

Figure 35.7 is the improved version of the graph, which has a higher data-to-ink ratio than does Figure 35.6 and is easier to read. I eliminated the Moiré vibration by using shades of gray and white and eliminated the legend by using labels directly on top of the appropriate areas of the graph. I removed the gridlines as well as the box around the graph and reduced the number of tick marks on the Y-axis to only the major ones. I shrank the type size of the years along

the X-axis and shifted to a different convention of representing the years ('94 instead of 1994) for all years but the first and last (another choice would have been to lay out the graph in landscape format so that the X-axis was longer). I also changed the figure caption to reflect the main story that the graph tells. Finally, I added author's contact information, the file name, and the dates of creation and last modification for the file.

A TRULY MISLEADING GRAPH

An example illustrating some important additional points about graph design appeared in the investor newspaper *Barron's* in September 2007, as shown in **Figure 35.8** (*Barron's* plotted the spot price of Saudi light crude oil, but the picture doesn't change much using the acquisition price for crude oil imported to US refineries and I happened to have those data handy, so that's what I plotted). The person interviewed by *Barron's* for that article wanted readers to look at the graph and conclude that history is replaying itself because the shape of those graphs are so similar. Unfortunately, that conclusion is not supported by these data.

As described in Chapter 33, this graph embodies a fundamental error by not correcting for inflation, but there are some other problems with it as well. First notice that the scales don't even cover the same ranges. The right hand scale, which applies to the data from the second period goes from $10 to $70 per barrel. The left hand scale, which applies to the first data series, ranges from $0 to $40 per barrel. So not only are the data in nominal dollars, which are not comparable, but they are scaled to different axes, compounding the confusion.

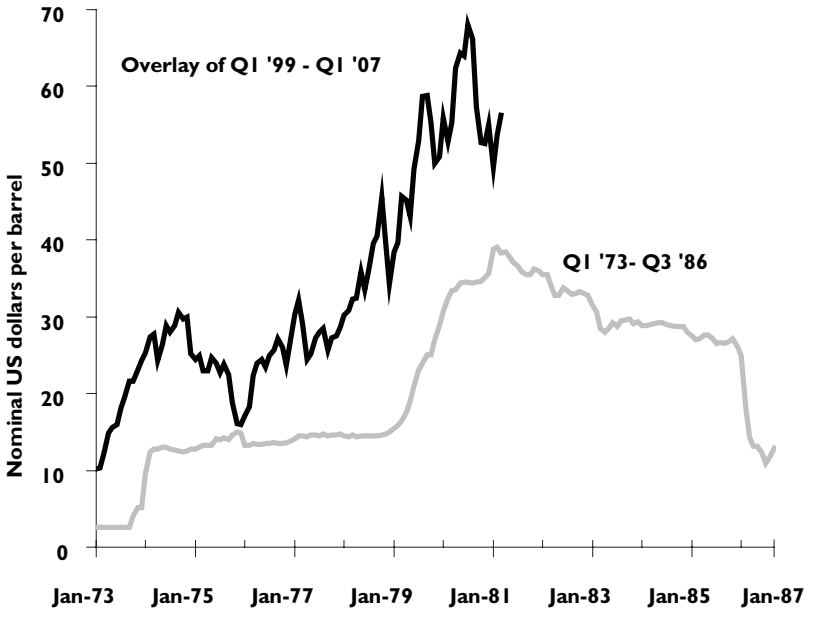
Using an axis that does not extend to zero without noting it for the reader is a problem that has been well known for more than five decades (with the publication of *How to Lie with Statistics* in 1954), but people still introduce this distortion with disturbing frequency. This technique helps to hide an important feature of the data in this graph. The curves look comparable, but the ratio of the highest price in each data series to the lowest shows that the magnitude of changes they represent isn't the same (even putting aside the error introduced by not correcting for inflation). The price increase in the first data series is about a factor of 15 in nominal terms from the first year to the year with the highest price, while it is only about a factor of seven for the second data series over a comparable time period.

So what happens if I plot the graph using the same axis for both series? **Figure 35.9** show the result, which makes the situation a little clearer.

FIGURE 35.8. Nominal refinery acquisition cost for imported crude oil to the US

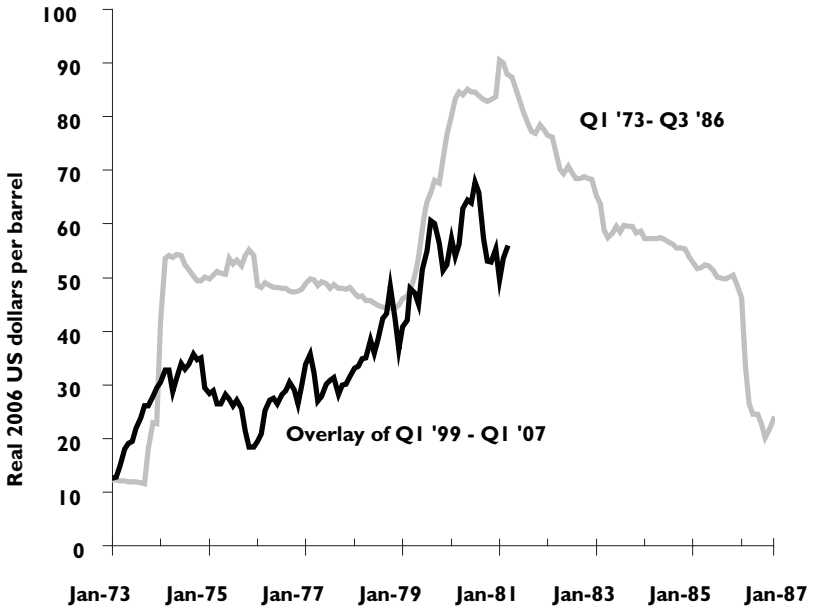
NOTE: Price data taken from the US Energy Information Administration <http://www.eia.gov/dnav/pet/pet_pri_rac2_dcu_nus_m.htm>.

FIGURE 35.9. Nominal refinery acquisition cost for imported crude oil to the US



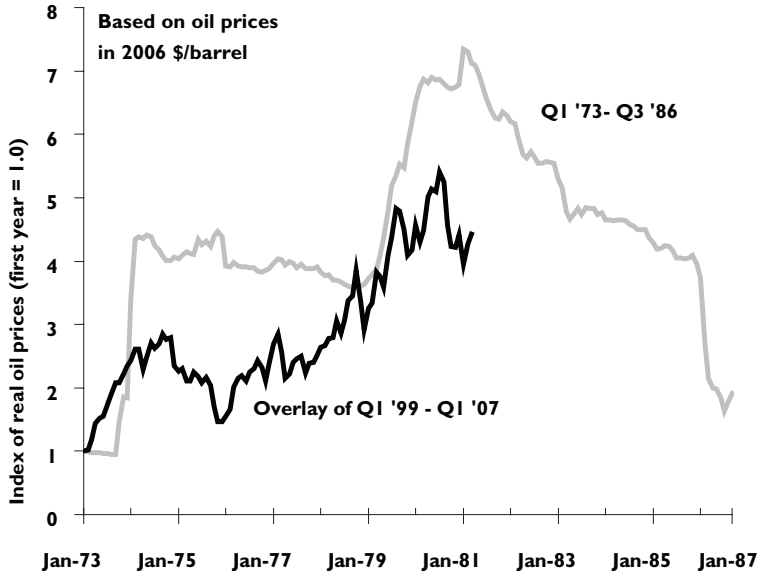
NOTE: Price data taken from the US Energy Information Administration <http://www.eia.gov/dnav/pet/pet_pri_rac2_dcu_nus_m.htm>.

FIGURE 35.10. Real refinery acquisition cost for imported crude oil to the US



NOTE: Price data taken from the US Energy Information Administration <http://www.eia.gov/dnav/pet/pet_pri_rac2_dcu_nus_m.htm>.

FIGURE 35.11. Real refinery acquisition cost for imported crude oil to the US



NOTE: Price data taken from the US Energy Information Administration <http://www.eia.gov/dnav/pet/pet_pri_rac2_dcu_nus_m.htm>.

Of course, comparing nominal dollars from different time periods doesn't make much sense. What does the graph look like after it is corrected to constant 2006 dollars? **Figure 35.10** shows the result.

If the purpose is to compare shapes of price trajectories over time, a clearer way to show these data would be as an index over time, as shown in **Figure 35.11**. Admittedly this graph looks pretty similar to Figure 35.10, but I believe it is a more straightforward way to make such comparisons.

These last two graphs tell a very different story. First, they show clearly that the magnitude of the price increase in real terms from the beginning to the highest price point for the two eight year periods is different (a factor of 7.4 for the earlier period and a factor of 5.4 for the later period). They also show that the shape of the price trajectories isn't nearly as similar as Figure 35.8 implied. In the first period, the real price of oil increased by a factor of more than four over a one-year period, then dropped only slightly over the next six years. After that period of modest price changes, the real price almost doubled over a two-year period. The graph representing the second period shows an increase in real terms of about a factor of three over a two year period, then a drop over one year that brought the price almost back to initial year levels. Prices then increased steadily over the next five years.

Comparing price trajectories over time is a difficult and thankless task, and one best left to the statisticians except in very rare cases. I would be hard pressed to argue that **Figure 35.11** shows anything other than two independent and unrelated price paths during periods of rising prices. It is only the distortions introduced by the use of nominal dollars and inconsistent axis scales that allowed the subject of the Barron's interview to argue that comparison of these two time series could yield insight into the course of future oil prices. You should not let such distortions color your judgment, your investment decisions, or your analysis.

CREATE STANDARDS FOR YOUR GRAPHICAL WORK

Stephen Few, author of S*how Me the Numbers: Designing Tables and Graphs to Enlighten,* suggests that business practitioners develop standards for graphical display. According to Few, such standards should

- have effective communication as their primary objective;
- evolve freely as conditions change or new insights are gained;
- be easy to learn;
- save time;
- be applied to the software that is used to produce tables and graphs in the form of specific guidelines and design templates.

In particular, he advocates that users of Microsoft Excel limit the types of charts they use to those that are truly effective and develop custom charts that improve upon Excel's default formats (which are typically appalling) to conform to standards that reflect their preferred design practices.

CONCLUSIONS

Tufte reminds us that rules should not be followed slavishly and that our focus should be on the key goal: "to give visual access to the subtle and the difficult."[80] Graphics should serve your arguments. If they don't support your analysis in a direct and understandable way, they should be ruthlessly deleted or drastically modified. As Tufte says, graphs and figures that don't contribute to the "revelation of the complex"[81] have no place in your work.

Read Tufte's books. Review his guidelines periodically, and apply them to your current work. Small changes in graphics design can vastly improve the effectiveness of your figures and graphs.

Visual representations of evidence should be governed by principles of reasoning about quantitative evidence. For information displays, design reasoning must correspond to scientific reasoning. Clear and precise seeing becomes as one with clear and precise thinking.

— **EDWARD TUFTE**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

36

CREATE GOOD TABLES

A good table is a work of art while a bad one is worse than useless. I have worked to refine my table-making skills over the years, and I am constantly amazed at the perfectly dreadful tables created by authors who should know better. Tables are the heart of a technical report, and care in their creation is one sign of a clear-thinking analyst. A little care can make your tables a resource that your readers keep as a reference for many years.

I follow several key principles in creating a table:

- *Make them stand alone*: Tables are often copied separately from the report from which they came. Make sure the reader can trace your table back to the source (or at least to the author) and that your calculations are explained clearly in the footnotes.
- *Create extensive footnotes*: This aspect of documentation is crucial. Footnotes to the table should explain the methodology, list the key sources, and describe how each source was used. Any intelligent person should be able to recreate the table from your footnotes. Tufte says that footnotes indicate "care and craft," and readers will be grateful when you create them.

  Of course, for tables appearing in more popular publications (like newspapers) the standards for footnote documentation are often lower, but even if the footnotes don't appear in the printed version, the person who created the table should still have the footnotes filed away in case there's a need to check the calculations later.
- *"Maximize the data-ink ratio," within reason*: Edward Tufte's guidance is just as relevant for tables as for graphs. Make sure that every part of the table conveys information. Avoid extraneous gridlines (only include them when they enhance reader comprehension). Use carefully chosen white space to separate and group results.

33
35

- *Some redundancy is good*: Unlike in graphs, where redundancy should generally be avoided, showing additional columns or rows for totals and subtotals is essential. These ostensibly redundant sums help the reader follow the calculation and assure that the table is internally consistent. The sums can also help in troubleshooting your own calculations. **27**

  Another important redundancy relates to percentages. If you have a column of percentages, label the column heading "% of total" AND put a percent symbol in every cell (e.g., 1.37%, not 1.37). Sometimes people make mistakes in their calculations and forget to divide by 100 to make a real percentage. By enforcing this redundancy, you guarantee that readers will interpret your results correctly and you'll save them time. Having the % symbol in every cell makes the meaning unambiguous.

  Finally, all numbers between zero and one should have a leading zero (e.g., 0.35, NOT .35). Copies become smudged and decimal points can be obscured. With a leading zero, there will never be ambiguity about the size of the number in question.

- *Enhance readability with font, style, and orientation choices*: In her *Non-Designer's Design Book*, Robin Williams demonstrates the principles of proximity, alignment, repetition, and contrast, and applies them to everyday design problems. Use spacing, justification (center, left, or right), grid lines, italics, and bold-facing to implement these principles within the table. Align decimal points within each column to add even more information.

  Choose a font that is easy on the eyes and doesn't deteriorate as a document is copied. Tufte recommends using the font used in telephone books (Bell Centennial) whenever you have tables that have high information density because the phone companies have researched this topic extensively, and there's no point in reinventing the wheel.

- *Enhance reader comprehension*: The single most important tool to increase usefulness of tabular data is to add rows or columns containing *indices* relative to a total or subtotal, which will show the relative importance of specific numbers relative to these totals (see the examples below). Most readers will calculate such indices themselves in their effort to understand the table. Do it for them, and they will thank you for it. Another **18** **27**

trick is to order table entries by size, thus allowing the reader the easily determine which are the most important. Another trick is to order table entries by size, thus allowing the reader to easily determine which are the most important.

- *Normalize data to facilitate comparisons:* It's often helpful to normalize results on a per capita or per GDP basis, so the reader can compare them to more familiar quantities. 27
- *Create the table first; then, write your text from the tables*: Creating the tables helps you think through the analysis and ensures that your logic is both correct and reproducible. Ask yourself questions about the analysis and check your work. If you can't explain it in writing, then you don't understand it. Clear tables and clear writing come from clear thinking. 34
- *Avoid spurious precision:* Only show the appropriate number of significant figures. Adding extra decimal places is distracting and pernicious. For example, results of calculations that use inputs reliable to one or two significant figures (e.g., 100 divided by 51) should not then be reported to five decimal places (1.93453). The correct value to use in this case is 2. The calculation is only as accurate as the least accurate input parameter (for more details on treating significant figures, see <http://www.chem.sc.edu/faculty/morgan/resources/sigfigs>).
- *Revise and edit*: Show your tables to colleagues before publication, and you will surely learn something. What is clear and obvious to you may not be so obvious to them. Give yourself time to revise and improve the tables.

A few specific examples will help illustrate some of these principles (A Microsoft Excel file [improvingtables2nded.xls] containing the original and improved tables can be downloaded at <http://www.numbersintoknowledge.com>).

The first example, in **Table 36.1**, is one contributed by Stephen Few, author of *Show Me the Numbers*. It illustrates some typical errors. First,

TABLE 36.1. A table with issues

YTD International Revenue

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Disk Drives	$93,993.00	$84,773.00	$88,833.00	$95,838.00	$93,874.00	$83,994.00	$541,305.00
Monitors	$87,413.00	$78,838.00	$82,614.00	$89,129.00	$873,020.00	$78,114.00	$1,289,128.00
Printers	$90,035.00	$2,120,400.00	$85,093.00	$91,803.00	$899,210.00	$80,457.00	$3,366,998.00
Computers	$92,736.00	$83,640.00	$87,645.00	$94,557.00	$92,619.00	$82,871.00	$534,068.00
Memory Sticks	$3,624,500.00	$77,785.00	$81,510.00	$87,938.00	$86,136.00	$77,070.00	$4,034,939.00
Sound Cards	$88,832.00	$80,118.00	$83,956.00	$90,576.00	$88,720.00	$79,382.00	$511,584.00
Video Cards	$82,614.00	$74,510.00	$78,079.00	$84,236.00	$82,509.00	$73,825.00	$475,773.00
RAM	$85,092.00	$76,745.00	$80,421.00	$86,763.00	$84,985.00	$76,040.00	$490,046.00
Scanners	$87,645.00	$79,048.00	$82,834.00	$89,366.00	$87,534.00	$78,321.00	$504,748.00
Input Devices	$90,275.00	$81,419.00	$85,319.00	$920,470.00	$90,160.00	$80,671.00	$1,348,314.00
Total	$4,423,135.00	$2,837,276.00	$836,304.00	$1,730,676.00	$2,478,767.00	$790,746.00	$13,096,903.00

TABLE 36.2. The table revised

YTD International Revenue (2008 USD)

Product	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Total*	*% of total*
Memory Sticks	3,624,500	77,785	81,510	87,938	86,136	77,070	$4,034,939	30.8%
Printers	90,035	2,120,400	85,093	91,803	899,210	80,457	$3,366,998	25.7%
Input Devices	90,275	81,419	85,319	920,470	90,160	80,671	$1,348,314	10.3%
Monitors	87,413	78,838	82,614	89,129	873,020	78,114	$1,289,128	9.8%
Disk Drives	93,993	84,773	88,833	95,838	93,874	83,994	$541,305	4.1%
Computers	92,736	83,640	87,645	94,557	92,619	82,871	$534,068	4.1%
Sound Cards	88,832	80,118	83,956	90,576	88,720	79,382	$511,584	3.9%
Scanners	87,645	79,048	82,834	89,366	87,534	78,321	$504,748	3.9%
RAM	85,092	76,745	80,421	86,763	84,985	76,040	$490,046	3.7%
Video Cards	82,614	74,510	78,079	84,236	82,509	73,825	$475,773	3.6%
Total	$4,423,135	$2,837,276	$836,304	$1,730,676	$2,478,767	$790,745	$13,096,903	100%

there's a great deal of non-data ink in this table. Second, there's excess precision in the numbers. Third, it's difficult for the reader to glean meaning from the data.

In the redesigned version (**Table 36.2**), the non-data ink has been removed, the dollar signs have been deleted (except for the totals), the numbers have been

TABLE 36.3. A table in one of my group's reports that needs revision

U.S. Shipments and Imports of Lamps, 1983-1994

Year	U.S. Shipments (millions of lamps)	Imports (millions of lamps)
1983	3615.9	560.9
1984	3723.4	748.7
1985	3472	862.7
1986	3421.3	920.6
1987	3399.4	999.8
1988	3510.2	1130.8
1989	3429.5	1024
1990	3318.5	1051
1991	3297.5	Data unavailable from Census Bureau
1992	3422.1	Data unavailable from Census Bureau
1993	3564.3	1372.6
1994	3563.3	1577.8

NOTE: "U.S. shipments" refers to total shipments by manufacturers located within the U.S., including units to be exported. Cold-cathode fluorescent lamps are excluded from the U.S. shipment data; Christmas tree lights are excluded from U.S. shipments as well as imports.
SOURCE: Census Bureau current industrial reports MQ36B, various years.

lined up by decimal points, the extraneous ".00" at the end of each number has been removed, and the categories have been ranked in decreasing order by the total. Finally, an additional column calculating the percent of the total has been added to aid reader comprehension.

The second example, shown in **Table 36.3** is taken from one of my former group's technical reports.[82] This table slipped through our quality control, but I'm making amends by dissecting it for this example. Note the extraneous gridlines and the lack of indices that would help the reader focus on the key results from the table.

Table 36.4 shows how I revised the table. I eliminated all but the essential gridlines and included indices relative to 1983 (which show how much shipments and imports grew from 1983 to 1994). I also added a column with imports normalized as the percentage of U.S. shipments, to help the reader

TABLE 36.4. Revised and improved version of Table 36.3

U.S. Shipments and Imports of Lamps, 1983-1994

	U.S. Shipments		*Imports*		
Year	Millions of lamps	Index 1983 = 1.00	Millions of lamps	Index 1983 = 1.00	Imports as a % of U.S. shipments
1983	3,616	*1.00*	561	*1.00*	16%
1984	3,723	*1.03*	749	*1.33*	20%
1985	3,472	*0.96*	863	*1.54*	25%
1986	3,421	*0.95*	921	*1.64*	27%
1987	3,399	*0.94*	1,000	*1.78*	29%
1988	3,510	*0.97*	1,131	*2.02*	32%
1989	3,430	*0.95*	1,024	*1.83*	30%
1990	3,319	*0.92*	1,051	*1.87*	32%
1991	3,298	*0.91*	NA	*NA*	NA
1992	3,422	*0.95*	NA	*NA*	NA
1993	3,564	*0.99*	1,373	*2.45*	39%
1994	3,563	*0.99*	1,578	*2.81*	44%
Average annual % change 1983–1994	-0.1%		9.9%		10.0%

NOTE: "U.S. shipments" refers to total shipments by manufacturers located within the U.S., including units to be exported. Cold-cathode fluorescent lamps are excluded from the U.S. shipment data; Christmas tree lights are excluded from U.S. shipments as well as imports.
NA = Not Available
SOURCE: U.S. Census Bureau current industrial reports MQ36B, various years
CONTACT: Jonathan Koomey, JGKoomey@lbl.gov. <http://enduse.lbl.gov>
FILENAME: shipmentsandimports.xls
DATE OF CREATION: February 1996
DATE OF LAST MODIFICATION: November 1997

understand how this important statistic changed over time. In addition, I included a row with the annual average percentage change from 1983 to 1994 because this summary statistic is helpful for understanding long-term trends in the data. Finally, I added information to the footnotes to identify the contact person, the file name, and the dates of creation and last modification for the file.

Table 36.5 shows a third example, taken from another technical report that my group worked on.[83] In this case, the client insisted on a particular format for the table, but **Table 36.6** shows how I'd change the table if it were up to me (it's

TABLE 36.5. A table that needs revision from another of our technical reports

Primary energy use in quads: 1973-1997

	1973	*1986*	*1990*	*1995*	*1997*
Buildings	24.1	26.9	29.4	32.1	33.7
Industry	31.5	26.6	32.1	34.5	32.6
Transportation	18.6	20.8	22.6	24.1	25.5
Total	74.2	74.3	84.1	90.7	91.8

SOURCE: Energy use estimates for 1973-95 come from EIA (1996a, Table 1.1, p. 39). Energy use estimates for 1997 come from EIA (1996c).

TABLE 36.6. Improved version of Table 36.5 for use in a summary report

Primary energy use in quads: 1973-1997

	1973	*1986*	*1990*	*1995*	*1997*	*Average annual change 1973-1997 % per year*
Buildings	24.1	26.9	29.4	32.1	33.7	1.4%
Industry	31.5	26.6	32.1	34.5	32.6	0.1%
Transportation	18.6	20.8	22.6	24.1	25.5	1.3%
Total	74.2	74.3	84.1	90.7	91.8	0.9%
Index (1973 = 1.00)	*1.00*	*1.00*	*1.13*	*1.22*	*1.24*	

SOURCE: Energy use estimates for 1973-95 come from EIA (1996a, Table 1.1, p. 39). Energy use estimates for 1997 come from EIA (1996c).
CONTACT: Jonathan Koomey, JGKoomey@lbl.gov. <http://enduse.lbl.gov>

not bad, but it could be better). To improve it, I added a column containing the average annual percentage change from 1973 through 1997. I also added a row with an index relative to 1973, which (for example) shows at a glance that primary energy use in the U.S. grew 24% over this period. Finally, I added contact information, so readers with questions know whom to approach. The more detailed information with the file name and date of creation and last modification makes sense for a technical table but would be excessive for a summary table.

TABLE 36.7. Improved version of Table 36.5 for use in a detailed technical report

Primary energy use: 1973-1997

	1973	*1986*	*1990*	*1995*	*1997*
	Quadrillion Btus of primary energy				
Buildings	24.1	26.9	29.4	32.1	33.7
Industry	31.5	26.6	32.1	34.5	32.6
Transportation	18.6	20.8	22.6	24.1	25.5
Total	74.2	74.3	84.1	90.7	91.8
	Sectoral energy as a percent of total				
Buildings	32%	36%	35%	35%	37%
Industry	42%	36%	38%	38%	36%
Transportation	25%	28%	27%	27%	28%
Total	100%	100%	100%	100%	100%
	Index 1973 = 1.00				
Buildings	1.00	1.12	1.22	1.33	1.40
Industry	1.00	0.84	1.02	1.10	1.03
Transportation	1.00	1.12	1.22	1.30	1.37
Total	1.00	1.00	1.13	1.22	1.24
	Average annual percentage change since 1973				
Buildings	0%	0.8%	1.2%	1.3%	1.4%
Industry	0%	-1.3%	0.1%	0.4%	0.1%
Transportation	0%	0.9%	1.2%	1.2%	1.3%
Total	0%	0.0%	0.7%	0.9%	0.9%

SOURCE: Energy use estimates for 1973-95 come from EIA (1996a, Table 1.1, p. 39). Energy use estimates for 1997 come from EIA (1996c).
CONTACT: Jonathan Koomey, JGKoomey@lbl.gov. <http://enduse.lbl.gov>
FILENAME: primaryenergyuse.xls
DATE OF CREATION: January 1999
DATE OF LAST MODIFICATION: December 1999

Table 36.5 appeared in a summary chapter of a larger report, so the recommended revisions are slightly different than if the table were to appear in a more detailed technical chapter (**Table 36.7** shows a version that would be appropriate in that situation). The same information on primary energy use in the different sectors is shown in that table, supplemented by several indices. First, the table shows energy use by sector expressed as a fraction of the total. Next, it shows indices relative to 1973, making growth over the analysis period readily apparent. The table also includes average annual percentage rates of change since 1973 for the sectors and the total. Finally, the additional detail about the file name and dates of creation and last modification are appropriate here.

Although some aspects of creating good tables are matters of style and preference, the principles described above should remain foremost in your mind, because they are valid for almost every table. Readers should treasure your tables for their accuracy and completeness. Sloppy tables sap your credibility—avoid them at all costs.

There are two ways of constructing a software design [or a table]: one way is to make it so simple that there are obviously no deficiencies; the other is to make it so complicated that there are no obvious deficiencies. — **C. A. R. HOARE**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

37

USE NUMBERS EFFECTIVELY IN ORAL PRESENTATIONS

Charles Osgood has a simple rule about using numbers in oral presentations: DON'T!

I disagree. You can use them sparingly, in carefully chosen settings. It is true, however, that most people can't absorb numbers and their implications without considerable hand-holding.

If you feel that numbers will enhance your argument, then summarize the few key numbers in graphic form. Only rarely use a table, and only then when it's a small one that contains the key information in large, readable type. If you have detailed data and calculations supporting your analysis, then give hard-copy handouts to your listeners. Summarize the conclusions from the results and focus on the conclusions and the story you are trying to tell, not on the numbers themselves.

Common doctrine is that data are dull. But as Edward Tufte points out, "if the statistics are boring, then you've got the wrong numbers" (or perhaps you are conveying them in a boring way). You should use numbers to tell a story or make an important point. If your audience is not yet convinced, then *you* are responsible for presenting the relevant information in a way that is engaging to them. Data are dull only when chosen poorly and presented badly.

32

Chip and Dan Heath, in their book *Made to Stick,* describe how some messages resonate and others fall flat. The ones that people remember are simple, unexpected, concrete, credible, emotional stories. They cite President Kennedy's goal of having a man walk on the moon within a decade and returning him safely to earth as the archetypal sticky idea. It was simple for people to understand, unexpected in its audacity, concrete and specific, credible because of the President's commitment of resources, and emotionally resonant with Americans' vision of themselves as intrepid explorers. Focus on using numbers to convey a few truly sticky ideas and your audience will thank you for it.

Make sure your numbers are correct! Nothing destroys your credibility faster than presenting numbers that are internally contradictory or obviously wrong. Once you lose credibility, it's nearly impossible to regain it during the course of a presentation. It won't matter how good the rest of your work is if you've presented bad data during the talk. So before you finalize your presentation, *dig into the numbers* and make sure they are both relevant and correct. Practice the talk before a friendly audience of colleagues, so they can catch any errors in advance.

27

Never display unreadable, densely packed tables or charts in presentations. I've seen numerous veteran professors do it, but it's insulting to the audience, and it's guaranteed to make your listeners tune out. These speakers are not trying to be disrespectful, they just assume that others find raw numbers as compelling as they do. Don't fall into this trap. You're the expert, and you should pre-digest the information and put it in an understandable form. It's not just common courtesy, it's also the best way to ensure that your main points are most effectively conveyed. If you want the audience to have the detailed numbers for reference, print and distribute well-documented handouts for your listeners so they can follow up at their leisure.

Gerald Jones, in his book *How to Lie with Charts,* gives guidelines on the size and kind of text to use in presentations and describes the conventions for measuring font size and spacing. He quotes the Society of Motion Picture and Television Engineers, who recommend that the minimum height of small letters used in on-screen text should be 2.5 percent of the height of the screen. Typical presentation screens are 1.5 meters (5 feet) in height, so the minimum font size in such a screen should be about 4 centimeters (1.5 inches). These guidelines don't just apply to text presentations. Make sure the labels on your tables and graphs are just as readable as your bullet points.

Another way to judge type size of an overhead projector slide is to place it at your feet. If you have good eyesight and can't read it while standing, the type is too small.

A successful speaker always thinks carefully about the following questions:

- Who will be in the audience?
- What is important to them?
- What are the three key messages you want people to learn from your talk? (With rare exceptions, that's all they will ever take away.)

32

Although Ross Perot's use of charts and tables did not earn him electoral victory in 1992, he demonstrated that a speaker who addresses issues the audience cares about can be effective and entertaining even if he uses more graphs per minute than in any television show in history.

COMMENTING ON THE PRESENTATIONS OF OTHERS

If you participate in conferences, you will occasionally be called upon (or may feel obliged) to comment on results presented in the oral presentation of another analyst. The same rules apply, but there are some subtleties.

29 Use back-of-the-envelope calculations to assess their analysis during the talk, but take extra care that your calculations are correct before speaking because you will likely be doing these calculations under time pressure.

Be polite! Not everyone accepts public criticism gracefully, even when it's gentle criticism. It's often difficult to engage in reasoned debate in a public forum, so in some cases it will be more fruitful to talk in private with the speaker after the presentation.

CONCLUSIONS

Effective presentations tell a story. Use carefully chosen numbers to tell that story in a compelling and succinct way. Consider what the audience cares about as you prepare your remarks, and make sure your talk gives your listeners the information they need and the means to find out more.

When speakers use a lot of numbers, the audience most always slumbers. — **CHARLES OSGOOD**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

38

USE THE INTERNET[84]

Anyone publishing analysis results in the 21st century had better take heed of the Internet. It is fast changing the way information is collected, processed, distributed, and used. Those who put the new medium to use will reap tremendous benefits; others ignore it at their peril.

FEATURES OF THE NET

In some ways, recent developments in electronic information distribution are a logical outgrowth and extension of the advent of mechanical type and printing presses five centuries ago. As James Burke points out in his book *Connections*, the trends reinforced by mechanical printing included rapid and accurate reproduction of texts, the specialization of knowledge and the responsibility of scholars for its accuracy, the first steps towards our current conception of intellectual property, and a sharp decline in the cost of reproducing and distributing information. The Internet accentuates each of these trends, but it also has its revolutionary aspects:

- *Near-zero marginal cost of reproduction and distribution:* Conventional printing vastly reduced the incremental costs of reproducing and distributing books, but the Internet reduces this cost to nearly zero. This shift from low marginal cost of reproduction to nearly zero cost is forcing another redefinition of our concept of intellectual property.

  Many companies are using the near-zero marginal cost of the Internet to reduce the cost of delivering customer support for their products. Frequently asked questions and product literature are now commonly made available electronically, which reduces the number of phone calls to customer support and therefore minimizes costs of fielding these inquiries. It also allows the technical support staff to focus on difficult problems with

fewer distractions. Shipping companies like Federal Express and UPS use the Internet for on-line package tracking, which vastly reduces their need for phone support.

- *Quicker publishing:* The barriers to publishing are now much lower because publishing materials on the net (and changing them) is relatively cheap, easy, and quick. The author Robin Williams summarizes Internet publishing as being "like a commercial print shop saying 'We'll print your recipe book this month for $30. Use as many colors and pages as you like. Change the recipes every day if you want, and we'll reprint it for you everyday, for free.'" This kind of flexibility does not exist in the world of conventional printing.

  On the other hand, lower barriers to entry in Internet publishing place a higher responsibility on each publisher. The Internet is a more freewheeling
16 place than conventional media sources, and it lacks the institutional credibility checks that are prevalent in newspapers, radio, television, and conventional publishing. It is therefore incumbent on each web publisher to establish and maintain its credibility in any way it can. Testimonials from established organizations with credibility can help with this effort, as can
33 clear and exhaustive documentation of sources and methodology.
39 - *Easier sharing of data:* One of the most interesting recent developments is the advent of web sites that make it easier to share data, including (by the end of 2007) Swivel, Freebase, Numberpedia.org, Data360, Graphwise, and Many Eyes. Each takes a slightly different approach to the problem, but they all share a common theme: giving users the tools to post, download, search, and (often) graph data. These sites haven't really solved the problem of assigning accurate credibility rankings to the data and their creators (it isn't sufficient to be 'the Youtube of data', for example—evaluating credibility is a lot more complicated than a five star ranking). They also have shown mixed success at promoting standardized data formats and complete documentation for data. But it's early days yet, and these sites are clearly the wave of the future in the data world.
- *Quicker review of technical material*: For analysis projects, making spreadsheets and draft reports available to readers on the Internet is a real boon. If readers have the spreadsheets containing the complete calculations, they can delve into the analysis much more easily and quickly than if they merely received a copy of the report. Full electronic access by end users speeds up the rate at which analysis can be reviewed, which reduces the time needed to create a final analysis.

- *Easier ordering and distribution:* Selling hard copy books and reports is easier and cheaper than in the days before the Internet. For certain kinds of books, at least, it is no longer necessary to have them stocked in local bookstores: people can find and order them on line, and receive them in a few days. This feature of the Internet eliminates a major hurdle for authors and small publishers.
- *Direct feedback from suppliers to consumers:* For manufactured products, many companies give an estimate of shipping time for a particular product on their web sites. The most sophisticated companies have a direct link between their computerized inventory management system and the web store that contains the product, which always reflects the current product supply. Never before have consumers known details about product supply and timing, and this capability gives them more flexibility in their purchasing decisions.
- *Direct feedback from consumers to suppliers:* Some companies are experimenting with using the Internet to allow their customers to supply specific information about company performance. The on-line auctioneer Ebay <http://www.ebay.com>, for example, has what it calls a "Feedback Forum," where buyers and sellers can comment on each other and on Ebay itself.
- *Indirect feedback from consumers to suppliers:* The Internet offers more and cheaper information about customers than ever before. Customers can specify the exact products they want, and this information is automatically collected in a database that can be "mined" by the company to learn about its customers. The company can also use information about who accesses its web site to create more targeted marketing campaigns. In addition, the company can track how people are using the search engine on the site to find out what kinds of questions people are trying to answer when they visit the site. This information can help the site developer improve the site but may also lead to ideas for new products to meet user needs.
- *Collaboration among users:* Well-designed collaborative web sites can allow customers and users to comment on the information distributed by the web publisher and to have those comments visible to others who visit the site. Such sites can also help users help each other by setting up discussion forums on topics related to the company's products. Interested users who have a specific question can visit such forums and have their questions answered by other interested customers, at little cost to the company itself beyond initially setting up the site. These user forums can also

be a fertile source of ideas for the company's product designers on improving existing products and creating new ones.

- *Access to information 24 hours per day:* Many information requests no longer need be made during business hours. Searching on the net can answer many questions at any time of the day or night. This flexibility reduces costs for all companies and saves users time, which benefits everyone. It is particularly beneficial when the company and the customer are separated by many time zones.
- *Universal searching:* The Internet allows access to vast storehouses of knowledge from all over the world and gives automated ways (search engines) to explore it. Never before has it been feasible to search such a vast network containing so much information. Finding essential data is often as much art as science, but experience with and study of search engine algorithms can help users find information quickly.
- *Easier and more widespread public access to technical information*: Technical information is often buried in impenetrable and obscure reports. Interactive links between the Internet and relational database management systems help those who possess detailed technical knowledge to make it useful to a wider audience.

  Lawrence Berkeley National Laboratory (LBNL), for example, has for decades been the preeminent center on energy use in homes, but much of the information LBNL generated never reached the general public until the advent of the World Wide Web. LBNL's Home Energy Saver web site <http://HomeEnergySaver.lbl.gov> was the first Internet-based home energy analysis tool; it embodies the technical expertise of dozens of people throughout LBNL. A user of this site has confidence that the tool accurately characterizes energy use in her home because of the expertise and credibility of those who created it.

  The Scorecard web site <http://www.scorecard.org> allows anyone to locate the sources of pollution in his own community merely by entering a zip code. It identifies the companies that produce the pollution and even lets users send a fax to the polluters complaining about it. The technical content of the site was created by people whose expertise in environmental issues is well known. They relied on publicly available data and compiled it in a form that any non-technical person could use to get immediate and understandable results. The information was available before the Scorecard team pulled it together, but no one had ever made it accessible in such an easy-to-use form.

The increasing availability of information about pollution, energy costs, political decisions, campaign contributions, and the spending habits of all types of institutions will have a larger and larger influence on our society during the next few decades. Once the data are compiled, they can be widely distributed at near-zero marginal cost. People need no longer take at face value the assurances of community leaders about topics of current interest. They can see and decide for themselves how to respond.

These features of the Internet present opportunities for analysts who are willing to seize them. The next section presents some specific advice for those who want to use the Internet to distribute data and analysis results with greatest effectiveness.

PUBLISHING ANALYSIS IN THE INTERNET AGE

There are three essential books on web publishing for analysts to read. The first is *The Non-Designer's Web Book*, by Robin Williams and John Tollett <http://www.peachpit.com/>, which contains reams of practical advice about making the Internet work for you. It gives information about color, file formats, typefaces, searching strategies, and graphics, and describes how Williams' design principles (from *The Non-Designer's Design Book*) should be applied to web sites. The second is *Philip and Alex's Guide to Web Publishing*, by Philip Greenspun <http://philip.greenspun.com/panda/>, which is a relatively advanced guide to the mechanics of web publishing, focusing heavily on techniques to enable collaboration between people and institutions. It is a bit dated now, but it is exemplary for its no-nonsense, sensible approach to dealing with complex issues, and it is also amusing. The third book is a more technical, textbook-style, and up-to-date follow-on to Greenspun's book, titled *Software Engineering for Internet Applications* by Anderson, Greenspun, and Grumet. I strongly recommend you read the first book cover to cover and browse the other two for relevant material.

In the section that follows, I assume familiarity with the basics of web publishing, so I can focus on key lessons of most relevance to analysts. For an introduction to the basics of the Internet, refer to Parts 1 and 2 in the Williams and Tollett book as well as the following page on Yahoo.com, which presents

resources for beginning web designers: <http://dir.yahoo.com/computers_and_internet/internet/beginner_s_guides/>. Here's my specific advice:

32 *Know your audience:* This is the first rule of all presentations, whether written or oral, and it applies just as well to web sites. Decide who will use your site, and make sure they can find what they want easily and quickly.

Focus on content, not graphics and fancy gimmicks: Many web designers are obsessed with graphics and the latest programming tricks. Including lots of graphics ensures that people using slower connections will spend a lot of time waiting. Relying on the latest web programming tricks ensures that people with older browsers won't be able to use your site (it also ensures that you will have to field technical support questions from users who can't get your web page to work on their computers).

Web sites should convey information, not display technical virtuosity. Determine the simplest approach to get your message out, and stick to it. Value
10 your user's time and design your site so that anyone can quickly and efficiently learn what you have to teach. One way to maintain this focus on the user's needs is to assign responsibility for content and usability to someone different from the web designer.

Do usability testing: Web sites are more like software than printed material. Even if you think you know your audience, you need to test the site in action. Greenspun recommends that you observe a few selected people using your site to accomplish specific tasks, which can help you "anticipate user questions and make sure your site answers them." Follow the guidance on "Involving users in the design process" found in the *Macintosh Human Interface Guidelines* (which is a time-tested source on how to make computers work well for humans). User testing is the best way to ensure that documentation is sufficient and that your user interface design is smooth as silk.

33 *Document like crazy:* Documentation is just as important for web pages as for books and hard-copy reports. Each page should have a "Date last modified" at the bottom as well as links that send email to the person who made it and to others responsible for the technical material described on the site. Don't forget

that web sites need to be backed up just like all other electronic information, both to safeguard the current version and to document past versions.

Use PDF: Portable Document Format (PDF) is a universally available and widely used format for distributing documents electronically. Readers can print PDF files that are exactly the same as the original report, which makes distributing reports electronically practically the same as mailing the reports (particularly as high-speed duplexing laser printers become more common). More sophisticated PDF files contain internal hypertext links, so you could jump from a listing in a table of contents directly to the relevant place in the document. PDF files can and should be made available for download from your web site.

Adobe Acrobat, one program used to create such files, is available for Windows and Macintosh computer platforms <http://www.adobe.com/products/acrobat/>. The Acrobat reader is free and it allows users of almost every type of computer, including Windows, Macintosh, Linux, and most forms of UNIX, to read and print PDF files <http://www.adobe.com/products/acrobat/readstep.html>.

Schemes for selling electronic content are still in nascent form, so there is a tension in the Internet world between free distribution (wide availability) and being rewarded for your efforts. To date, no proposed methods for addressing these issues have caught on. One clever approach was championed by the now defunct company called Softlock.com. Its system allowed anyone to view and email a sample of a given PDF document. To see the full document, the reader had to go to the Softlock.com web site and pay for access. Once the fee was paid, the full version of the PDF file was unlocked for use on the user's computer. If the file was emailed or copied to another computer, it reverted back to the sample so that the recipient of the file could evaluate it and decide whether it was worth buying. The creator of the file could specify the level of access that purchase of the file grants (read only, read and print only, read/print/copy text, etc.). Unfortunately, the technology did not achieve market success (which is an oft-repeated story for such schemes).

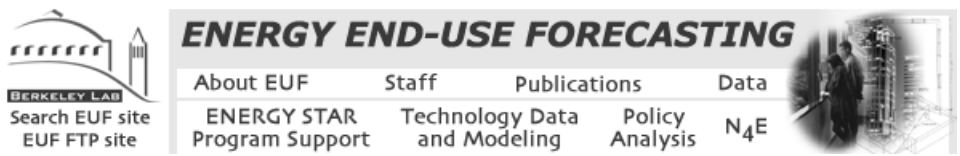
Assign one web page per publication: Assign every new publication a unique web page, and write the URL on the title page of the document. This web page should include related data, analysis spreadsheets, related links to other people's sites, errata discovered after publication, and (in many cases) downloadable PDF files of the report itself. Owners of the hard-copy report know that they

can always go to that page to get further info. They can also direct other interested people to that web page to avoid having to make a copy themselves. This approach requires a file structure that does not change after the site is designed, so make sure you've settled on a sensible structure before widely publicizing your site (this is good advice in any case).

Make data and calculations available: For web pages from which users will access data files, post downloadable spreadsheet files with plenty of description of the files on the web site, including file format, name, version number, date last modified, and size. The link to download a particular file might then look like this: *downloadme.xls* (150k), version 2, in Microsoft Excel 97/98 format (last modified 31 October 1999). This information can help a user decide whether to download the file. That user may have a slow network connection, for example, and may not want to download large files, or may not have Microsoft Excel. By giving users the file size and format, they can make the decision that is most
33 appropriate for each situation. Be sure to document your calculations and data sources in the file itself so that you avoid readers badgering you with questions that should have been answered by your footnotes in the first place.

From the perspective of making data available for calculations, it is almost always better to have a downloadable file than an HTML page that you must save as text (of course the HTML page is easy and quick to read). If you have the time, it's best to make both available to your on-line readers, to serve both the need for quick access to some of the data and the need for easy spreadsheet
27 access to all the data. Cleaning the stateline.org data would have been a lot quicker with a downloadable file.

Use repetition to good advantage: Develop a uniform format for the web site's
35 pages. This approach relies on the same principle that makes Tufte's "small multiples" so effective in conveying information: once readers understand the format, they need not re-examine it each time they access the information anew. My old group's site at LBNL used the same navigation toolbar at the top of every page. It looked like this:



The site also had a text toolbar at the bottom of every page that replicated many of the functions of the graphical toolbar, for people with slow connections who browse the web with graphics turned off.

The text toolbar looked like this:

About EUF | Staff | Publications | Project Areas | Data | End-Use FTP Site | Search

Wherever readers found themselves on the site they always had these toolbars available and could access the information they needed quickly and effectively.

The same principle applies to the content of the pages themselves. That site also had unique pages for each distinct project and each project page had consistent headings:

> *Project Description*, which gave a few paragraphs describing the work and an occasional figure to illustrate important results;
>
> *Project Staff*, which was a list of those who worked on the project with a direct link to each person's biography page and contact info;
>
> *Key Data*, which included downloadable files that contain inputs and results for the analysis;
>
> *Publications*, with full references to the reports as well as links so that the complete PDF files could be downloaded by the reader;
>
> *Other Resources*, which included links to external sites related to the project;
>
> *Sponsors,* which listed the people and institutions responsible for funding the project.

Readers knew that the project pages all had these headings; this familiarity made it easier to find relevant information quickly.

Pack your site with carefully chosen text: Search engines look for text, so the more text on your site, the better (this is another reason to focus on textual content and not on graphics). Of course, your pages should attract readers and help them find the information they need, so follow Williams' advice about proximity, alignment, and contrast in this effort. Don't overwhelm readers, but do put as much content on your site as makes sense. Greenspun suggests that you consider purchasing the rights to text that is related to the purpose of your site and posting it. This approach is cheaper and more likely to attract visitors than purchasing on-line advertisements on other sites (anyone using a search engine

to find information on a topic related to that of your site will have many links to your site pop up from their search query).

If you want to attract search engines to a page that is predominantly graphical, HTML allows programmers to include "meta-tags" that contain key words summarizing the purpose of the page and site. Another trick is to put text on the bottom of that page that is white on white background. It won't appear to those viewing the page with a browser, but the search engine will use that text to index the page.

Enable collaboration: One of the most intriguing and powerful aspects of Internet publishing is the potential for on-line collaboration. Web sites that encourage readers to add comments, ask questions, and answer queries become a treasure trove of information, not just for the users of the site, but for the publisher as well.

Another kind of collaboration is enabled by the use of "Project sites" <http://www.secretsites.com/>, which contain the project's prospective schedule, an email trail, the historical chronology of project developments, and the relevant electronic documents. Creating such a password-protected site makes the key information available to all participants in the project and allows them to work together more effectively. For one quick and easy way to set up such a site, go to <http://www.basecamphq.com>.

Still another form of web-enabled collaboration taps the innovative spirit and subject knowledge of a company's employees. As documented in a Harvard Business School Case, a Rhode Island–based firm called Rite-Solutions <http://www.ritesolutions.com> pioneered this idea in January 2005 by creating its own internal marketplace for ideas, called "Mutual Fun". Ideas listed as stocks on the exchange could involve saving money for the company, using existing technologies to create new products, or striking out in entirely new directions. To get a stock listed, an employee with an idea would need to enlist a sponsor (one of 40 high level managers approved by the company's leaders) who would help foster the idea. Each employee was allocated $10,000 in virtual "money" that could be allocated to any stocks on the Mutual Fun exchange. Then the fun began. The best ideas quickly rose to the top as their stock prices increased. People who wanted to help improve an idea had a formal online process for submitting their thoughts, which helped make their interactions with the creators of the stock as productive way as they could be.

Mutual Fun worked so well that Rite Solutions now licenses the technology to other companies for their own internal use.

Interact with users: Greenspun says "User feedback might be annoying at times, but if you want to be insulated from your readers, why publish on the Web at all?" The Internet allows closer connection between customers and businesses than previously possible. Use those connections to your advantage, and learn about your customers to serve them better. Collaborative services can help in this learning process, as can analysis of server logs and direct email interaction. Give users the ability to comment on your site and your products, and you are certain to learn things you could not have learned on your own.

Set up an internal search engine and site index: A search engine devoted to your site will save your readers time and frustration. Greenspun suggests that you program it to send you email if a search returns no items or is repeated several times: "As soon as two users ask the same question, you should beef up your site to answer it unattended." Thus, the search engine can help you improve your site, assist your users, and teach you about their needs all at the same time. A site index, which lists keywords with links to the related part of the site, will also help readers find the information they need.

Get listed in external search engines: Make sure you're registered with the major search engines and directories because these services are likely to be your most important source of web traffic (there are companies that do this registration for you, for free or for a nominal fee). Such registration insures that people searching for the information you have will be able to find you quickly. Some of the most important search engines (in alphabetical order) are Altavista, Excite, Google, HotBot, Infoseek, Lycos, Webcrawler, and Yahoo. For the latest information on this process, check out the advice available at <http://www.searchenginewatch.com>. Williams and Tollett point out that some search engines specialize in certain types of technical information. If there's one for your specialty, then make sure you register with that one as well.

Use printed material when appropriate: One downside to the Internet is that reading text on a computer screen is still not as convenient or as pleasant as reading a book. Printed documents can be read anywhere and have a clarity and

permanence that electronic media cannot yet match. Until the technology for electronic books becomes better developed, paper books and reports will remain important (though *distribution* of some of these documents will surely be replaced by PDF files that can be electronically distributed).

Link to other sites: Reports that only reference the author's own work rightfully encourage skepticism, and web sites that don't refer to other kindred sites deserve similar treatment. Add value to your site by helping readers find other sites that provide complementary information and data. Access to comprehensive lists of related links is one of the reasons web surfers return to a site again and again.

Bring in the big guns: When you have a lot of data that you want to make available, call in the experts on database-backed web sites. Greenspun describes the trials, tribulations, and triumphs of the people who make these complex sites happen. The Home Energy Saver and Scorecard web sites are two good examples of how this kind of programming can make complex data and analysis results accessible to a wide audience. These sites can be quite expensive to develop and maintain but can be indispensable for certain applications.

CONCLUSIONS

Seize the opportunities the Internet presents. After defining your audience, create content that matters to them and make it available on a well-designed and well-tested site. Avoid fancy graphics because they consume bandwidth and contribute little to getting your message across. Make sure your site is accessible to people using modems, older browsers, and all major operating systems. Set up comment pages and discussion forums to help your readers teach you and each other. Finally, remember that the Internet is just another tool and that it's not *always* the right tool for the job. Non-electronic publishing still has its place.

When I am working on a problem I never think about beauty. I only think about how to solve the problem. But when I have finished, if the solution is not beautiful, I know it is wrong.

— **BUCKMINSTER FULLER**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CHAPTER

39

SHARE AND SHARE ALIKE

One of the biggest problems with data is that they often aren't linked to the critical information needed to make the data useful, and thus are "disembodied". For example, a downloaded spreadsheet will usually not contain a complete citation for the data itself or a description of the methods and sources used to compile it.

Disembodied data are also commonplace on the World Wide Web. Consider a fictional address: 54321 Broadway Avenue, Sometown, NY 12345. If it appeared on a normal web page, search engines or software agents could only rely on the pattern of characters to match against a search query, using "natural language processing". They could not know for sure that 12345 is a zip code, or that 54321 Broadway Avenue is a street address.

In both cases, the data lack *metadata,* described by Wikipedia as "data about data" <http://en.wikipedia.org/wiki/Metadata>. The American Library Association more precisely defines it as "structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities." Most computer users are familiar with the vCard format for address book data, which allows software to import and export contact information. vCards associate all parts of addresses with correct metadata (for example, all zip codes are correctly labeled as such).

Lack of metadata prevents software from making more intelligent use of data on the web. To solve this problem, Tim Berners-Lee and other web pioneers have developed the "Semantic Web". This innovation would allow web page creators to associate metadata with important data types, and to make the characteristics of and the relationships between such metadata universally accessible to users of the web. In practice this would allow the web to continue to

grow rapidly at the same time enhancing the abilities of software tools to access and use relevant data. Berners-Lee summarizes his big-picture view of the potential as follows:

> Human endeavor is caught in an eternal tension between the effectiveness of small groups acting independently and the need to mesh with the wider community. A small group can innovate rapidly and efficiently, but this produces a subculture whose concepts are not understood by others. Coordinating actions across a large group, however, is painfully slow and takes an enormous amount of communication. The world works across the spectrum between these extremes, with a tendency to start small—from the personal idea—and move toward a wider understanding over time.
>
> An essential process is the joining together of subcultures when a wider common language is needed. Often two groups independently develop very similar concepts, and describing the relation between them brings great benefits. Like a Finnish-English dictionary, or a weights-and-measures conversion table, the relations allow communication and collaboration even when the commonality of concept has not (yet) led to a commonality of terms.
>
> The Semantic Web, in naming every concept simply by a Uniform Resource Identifier (URI), lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together.

Inadequate metadata also presents a challenge to good analytical practice that is in some ways more complicated than that for the web itself. Data shared among analysts is even more diverse in its structure and format than the bulk of those posted on the web. Full documentation requires deep conceptual understanding and a level of attention to detail that is rare. And gleaning meaning from data often requires expert contextual knowledge that is time consuming and difficult to summarize for those who are not familiar with the practices and terminology of the relevant expert community.

In spite of those challenges, the potential benefits of more efficient, effective, and accurate data sharing could be substantial, and that has led to a profusion of new web sites devoted to sharing data and helping users graph it. Unfortunately, the current sites have at best a mixed record of capturing these

benefits, in part because they underestimate the importance of structured data. This chapter reviews the promise and pitfalls of recent attempts to enhance data sharing, and describes how they might be improved in the future.

In this chapter, I'm focused specifically on what I call *public data,* which are freely available to anyone. These data include those available from government sources, but are not limited to those sources. For example, the income statement for any publicly traded company for 2007 contains public data—anyone who wants to know about profits and losses for that company in that year can request it. Society has determined, for reasons of transparency and accountability, that those data should be calculated in a standardized way and made available for public scrutiny (the accounting fraud associated with the Enron and Worldcom debacles notwithstanding). I am not specifically focused in this chapter on proprietary data internal to companies, although much of the discussion here will be relevant to those data as well.

THE PROMISE OF DATA SHARING

In creating data for use by others, there are two basic approaches: the "wiki" model (where anyone can contribute information or change documents, with the document becoming the sum total of the contributions of many members of the community) and the "expert" model (where a recognized expert on a particular topic creates a document or compiles data and the community trusts the compilation because of that person's expert knowledge). The struggle between these two philosophies has existed since the dawn of the World Wide Web, and it is unlikely to be resolved anytime soon. Wikipedia, of course, has championed the wiki model, while Google has focused recently on the expert model by introducing "Knol".

It is clear that for some kinds of information the pure wiki model works well, but for data sharing, the expert model has advantages. Numeracy skills matter, and competent experts are better able to compile data in a way that is meaningful and useful. People who are not experienced in numerical issues will make mistakes that a more numerate person generally will avoid, and those mistakes will propagate quickly in an online data sharing community. Of course, some parts of the wiki model related to on-line communities, like user comments on

data (which represent a type of informal peer review), will still be important when using the expert model, but the central role of experts in creating and posting the data would not be diminished by the use of collaborative web tools.

Assume for a moment the existence of data sharing web sites that enforce rigorous documentation of sources and methods, track the change history for the data, attach units to all numbers, and track multifaceted rankings of the reliability of the data itself, as well as the reputation of its creators and the institutions where they work and publish. If sites of that type existed, what benefits would they yield for users? Three major areas immediately come to mind: quality control, ease of finding and using information, and visual discovery.

Quality control

Data quality is a multidimensional concept encompassing completeness, accuracy, and contextual assessments of the relevance and credibility of data. Web sites sharing structured data could easily enforce basic measures of data quality. Incomplete sources, inadequate explanatory documentation, missing values, and unlabeled units would negatively affect the confidence users have in its underlying quality.

Improved quality control would make analysis more rapid and accurate and minimize some of the pitfalls described earlier in this book. Cleaning the data, finding complete references, and understanding how the data were derived would be all be made easier. All analysts would benefit.

Ease of finding and using information

Web sites that facilitate data sharing would allow searches to be much more efficient. In particular, full references and descriptions of how the data were created would provide quicker and more accurate searches. Keyword categories developed specifically for certain types of data would also improve matters. And once users found information, they would have the confidence and quality rankings to help them assess it, plus the full documentation to help them reference it correctly in their work.

Visual discovery

Sharing of structured data would allow software to more easily and accurately graph it. For example, having the units attached to each data series would

allow software to identify conflicts and prevent data series with incompatible units from being plotted on the same graph. While having the ability to graph data on-line is convenient, making good graphs is already possible using existing desktop software—it is *sharing* of structured data that is most important at this stage (a point that is unfortunately often neglected by current data sharing sites).

CURRENT DATA SHARING SITES

As of February 2008, I've been able to locate the following web sites devoted to data sharing (in alphabetical order):

Data 360 <http://www.data360.org>: This site allows easy posting of data as well as graphing of those data. It explicitly focuses on sharing of public data, but also allows companies to post their own data for a fee. It seems to tilt towards the expert model—I wasn't able to find any way to comment on or change data sets posted by others, although the site hints that people other than the creator can update data.

Freebase <http://www.freebase.com>: Freebase, a product of Metaweb technologies, is described by its developers as an "open database of the world's information". It is a well-funded startup company, and has developed a sophisticated way of using on-line databases that incorporates characteristics of both the wiki and expert models. It's still in Alpha as of this writing, but it shows great promise. It doesn't allow graphing of data within the site, but gives users the ability to grab any data from Freebase and use it in other applications.

Google Base <http://base.google.com>: Google base allows you to post data in more than a dozen predefined item formats (called "types"), including job advertisements, vacation rentals, recipes, products for sale, real estate, events, or personal ads (among others). You can also define your own data type, assuming it's not a particularly complex one. If your data can be structured in a straightforward way, Google Base will make it easy to post it (and by extension, easy for people to find your information using Google's search engine). You can even upload data continuously or in bulk. There is no graphing of data in Google Base, and no way for a user to correct or modify data posted by others.

Graphwise <http://www.graphwise.com>, defunct as of March 2013: Graphwise was the newest of those described here. This site claimed to be "primarily a search engine" that finds data on the web and automatically creates "meaningful plots of that information". As Stephen Few pointed out in his blog <http://www.perceptualedge.com/blog/>, that effort is fraught with pitfalls, and the site never really solved those issues . The site was less focused on users graphing their own data. For example, I was able to easily upload a small excel table into the site when it was operating, but was unable to figure out how to make a graph from it.

Many Eyes <http://www.many-eyes.com>: The purpose of Many Eyes is data visualization. Its data sharing abilities are somewhat limited, but its tools for graphical display of data are the most sophisticated of all these sites. As with all graphing tools, it is possible to make nonsensical graphs from any data, but the graphing tools of Many Eyes are better thought out than others I've used.

Numberpedia <http://www.numberpedia.org>, defunct as of March 2013: Numberpedia focused exclusively on what I call "factoids"—individual statistics that tell a story. It used the wiki approach and did allow units to be attached to each statistic. There was no obvious way to post an array of numbers (like a time series), and the site said that attaching the source to each number was optional, which is inconsistent with my principles for documentation.

Swivel <http://www.swivel.com>, defunct as of March 2013: Swivel was sophisticated in enabling collaborative interaction. It focused mostly on graphing and the use of collaborative tools but it also gave several convenient ways to import tables (arrays) of data, including a toolbar for Microsoft Excel to automate the process. It tilted more towards the expert model (going so far as to have "official data sources" labeled on the site), although it did allow users to comment on data sets that they did not create.

These sites apply social networking and on-line collaboration tools to enable data sharing. They share a few common attributes:

- All allow posting/sharing of data.
- All force some structure on the data, but not much
- Most allow users to comment on the data.

- Some allow users to rank or rate data (although what criteria they are using to rank is usually unclear).
- Some allow easy download of data to Excel or other easily readable formats
- Some allow graphing of data
- None enforce adequate documentation

With the exception of Google Base, which is a somewhat different animal than the others, these sites have real quality control issues. In particular, I saw many instances of data with positive or negative ratings but no way to interpret what those ratings mean, data without any documentation, data series with questionable entries, graphs with nonsensical axes, graphs plotting clearly unrelated data on the same chart, and graphs with incorrectly labeled units.

ENABLING DATA SHARING IS THE KEY

The designers of the sites described above have applied the social networking and collaborative models that have worked well for YouTube and Wikipedia without giving careful thought to what would actually work best for data. The scientific community has taken a different approach, in which the technical, legal, and logistical issues of data sharing are front and center. Creative Commons, for example, has addressed ambiguities in current law that impede sharing of databases by creating an open access data protocol <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>, as part of their "Open Data Commons" project <http://www.opendatacommons.org/>. Creative Commons correctly (in my view) understands that core issues of data sharing (including legal issues and technical/practical issues about structured data) need to be resolved before data sharing web sites can reach their full potential. This realization is as true for more popular data sharing sites as it is for their scientific counterparts.

CONCLUSIONS

Ideal data sharing sites would allow the community to rank the credibility and quality of data. They would make it easier to find data and to design software

tools to graph those data in a way that would yield insights, accelerating both technological and intellectual innovation. There are clear advantages to computer-aided sharing of structured data, but at this writing existing data sharing web sites fall far short of reaching this promise.

Whenever I found out anything remarkable, I have thought it my duty to put down my discovery on paper, so that all ingenious people might be informed thereof.

— **ANTONIE VAN LEEUWENHOEK (1632–1723)**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

ACKNOWLEDGMENTS

I would like to thank the following people for their comments and encouragement throughout the process of writing this book: Todd Baldwin, Eric Bergman, Marcela Bilek, Jeanne Briskin, Erik Brynjolfsson, Johanna Buurman, Chris Calwell, Sheryl Carter, Cynthia Chaffee, Eva Cohen, Ken Conca, Laura Driussi, Carol Emert, Chris Ertel, Marta Feldmesser, Maria Fink, Karina Garbesi, Steve Greenberg, John Harte, Dianne Hawk, Rick Heede, Ephraim Heller, Karen Heller, Chris Koomey, Gregory Koomey, Richard Koomey, Skip Laitner, Tod Loofbourrow, Amory Lovins, Nathan Martin, Alan Meier, Evan Mills, Marja Mogk, Amey Moot, Mark Morland, Laura Nelson, Laura Newcomer, Andrew Nicholls, Mary Orland, Joe Romm, Art Rosenfeld, Erika Rosenthal, Marc Ross, Suzanne Samuel, Polly Shaw, Lene Sørensen, Philip Sternberg, and Kim Taylor.

Special thanks to R. Cooper Richey at Centaurus Energy and Michele Blazek at AT&T, whose careful reading vastly improved this manuscript.

I would also like to express my gratitude to those friends and colleagues who were most instrumental in teaching me the "tricks of the trade" over the years, including Florentin Krause, Mark Levine, Jim McMahon, Art Rosenfeld, John Harte, John Holdren, Barbara Barkovich, Bill Ahern, and Amory Lovins.

Wylie O'Sullivan and Nan Wishner performed careful copy editing of the manuscript, and Veronica Oliva alerted me to possible issues related to permissions for copyrighted material cited within (though I made all decisions about the use of copyrighted material and permissions for that material). Susanna Tadlock supervised the production and printing of the book (a task for which I owe her special thanks), Sandy Drooker is responsible for its lovely design, and Dave Peattie and Bea Hartman did a masterful job of typesetting.

Tom Chen's graphics grace the cover and a couple of the chapters. He deserves my thanks for creating art that really works with the book.

Finally, I would like to dedicate the second edition of this book to my beloved Melissa, who by her very existence has made my world brighter and more wonderful.

© 2001 by Tom Chen. All rights reserved.

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

CONCLUSION

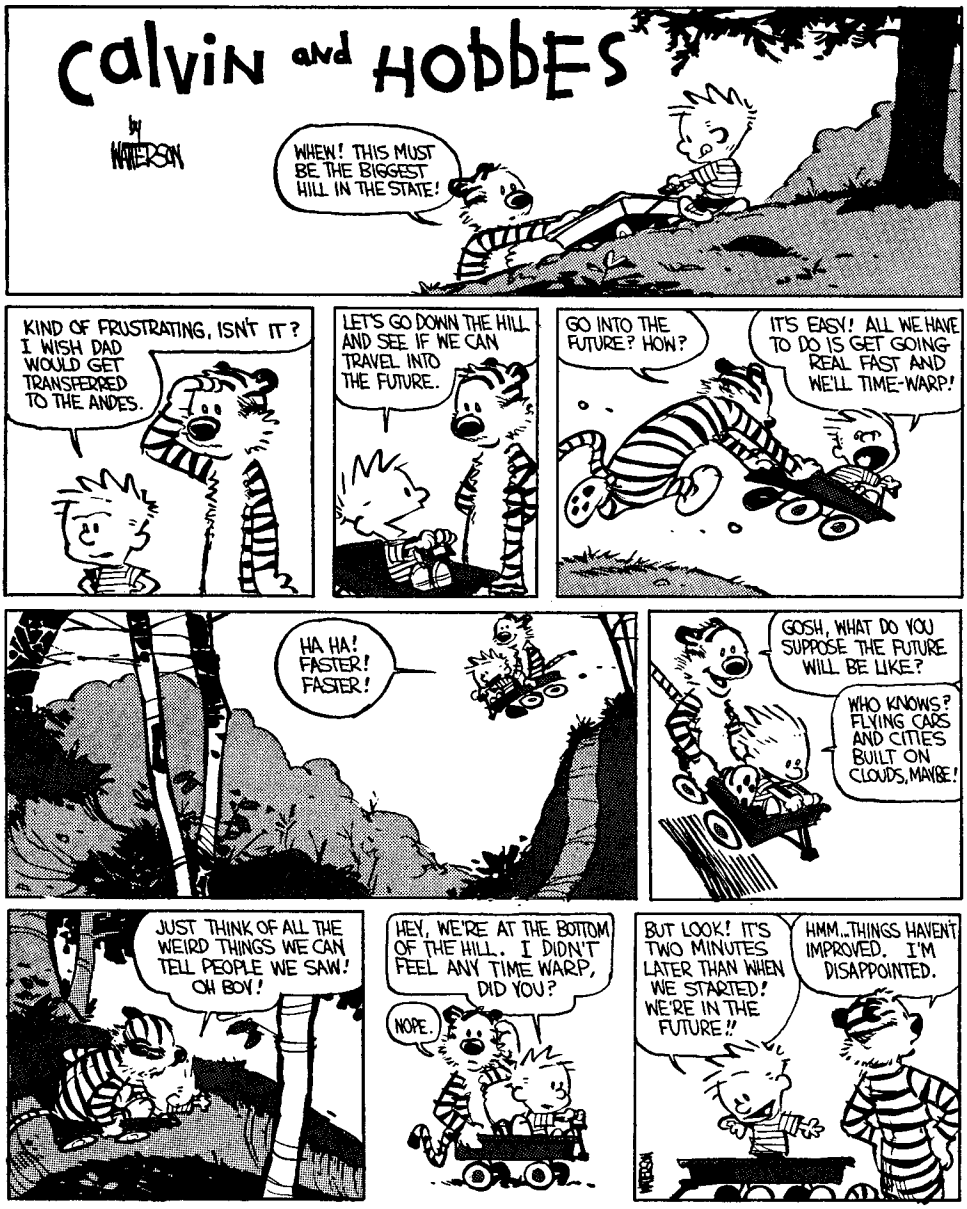
CREATING THE FUTURE

Like the butterfly's wing beat that ultimately leads to a hurricane, seemingly trivial actions can have consequences that reverberate through generations. With every action, with every day we live, we create the future. Of course, forces beyond human control also have influence, but it is how our choices relate to these external events that determine the outcome.

Mastering the art of problem solving is one important way to ensure that the future you create is a hopeful one. Analysis helps you make choices that are consonant with your personal values, consistent with your ideology, and reflective of the realities imposed by external forces. Whether you are a businessperson, scientist, athlete, or accountant, analyzing your choices in a systematic way will lead to a better life.

Of primary importance from this book are the following lessons:

- *Don't be intimidated by anyone:* The key issues surrounding all but the most technical topics can be understood by any intelligent and sufficiently diligent person. **2**
- *Be a critical thinker:* Dissect your own arguments and those of others. Identify the premises and the main conclusion of each argument. Make sure the premises are acceptable, relevant, and adequate to support the conclusions. Finally, search for missing arguments and for counter arguments. **11**
- *Don't confuse what's countable with what really counts:* Many of the most important things in life can't be quantified, so don't focus just on the numbers–they aren't everything. **12**
- *Get organized:* Clean up your office. Establish a filing system that allows you to find what you need when you need it. Buy the key books in your field, and make them accessible. Above all, remember that organization often beats brilliance! **6** **7**
- *Question authority:* Maintain a healthy skepticism of other people's analysis, and ask questions until you understand their points. Clearly dis- **14**


CALVIN and HOBBES
by WATTERSON
WHEW! THIS MUST BE THE BIGGEST HILL IN THE STATE!
KIND OF FRUSTRATING, ISN'T IT? I WISH DAD WOULD GET TRANSFERRED TO THE ANDES.
LET'S GO DOWN THE HILL AND SEE IF WE CAN TRAVEL INTO THE FUTURE.
GO INTO THE FUTURE? HOW?
IT'S EASY! ALL WE HAVE TO DO IS GET GOING REAL FAST AND WE'LL TIME-WARP!
HA HA! FASTER! FASTER!
GOSH, WHAT DO YOU SUPPOSE THE FUTURE WILL BE LIKE?
WHO KNOWS? FLYING CARS AND CITIES BUILT ON CLOUDS, MAYBE!
JUST THINK OF ALL THE WEIRD THINGS WE CAN TELL PEOPLE WE SAW! OH BOY!
HEY, WE'RE AT THE BOTTOM OF THE HILL. I DIDN'T FEEL ANY TIME WARP, DID YOU?
NOPE.
BUT LOOK! IT'S TWO MINUTES LATER THAN WHEN WE STARTED! WE'RE IN THE FUTURE!!
HMM...THINGS HAVEN'T IMPROVED. I'M DISAPPOINTED.

CALVIN AND HOBBES © 1987 Watterson. Dist. by UNIVERSAL PRESS SYNDICATE.
Reprinted with permission. All rights reserved.

19 tinguish facts from value judgments, and make sure that others do so too.
16 Don't believe everything you read.

27 • *Dig into the numbers:* Look for internal contradictions, large differences, obvious trends, and cognitive dissonance. Compare results to independent sources. Take ratios of numbers, and make sure they relate in a predictable way.

- *Focus on the essential:* In creating your analysis, focus on designing consistent comparisons, developing credible scenarios, and using simple and understandable models to illustrate the key issues. Avoid complex computer tools unless absolutely necessary. **25** **26** **28**
- *Document, document, document:* Documentation creates a trail for you and others to follow after the analysis is done. It is essential to any intellectual work but is oft neglected. By making it a priority, you ensure that your intellectual contribution will remain important and useful for years to come. **33**
- *Use the Internet*: Rely on the Internet to conduct your research, share your data, and make your publications widely available. The old ways of publishing and disseminating data are rapidly being swept away by electronic media, so use these media to full advantage. **38** **39**
- *Remember that others don't care as much about your work as you do:* It's your responsibility to know your audience, address their concerns, and explain your work in an understandable way. Make your figures clear and compelling, your tables well documented, and your reports so valuable that your colleagues will treasure them for years. **32** **35** **33** **36**
- *Synthesis follows analysis:* Combining analysis results from different fields in creative ways often leads to new insights. Such synthesis is one powerful way to increase your understanding of the world around you. Don't underestimate its importance.

Finally, the integrity of your calculations and the clarity of their presentation are both critical to your analytical success, but even more important is your focus on the issues that really matter. Too many people expend their effort on topics of only marginal importance. Your time is your life, so make it count! **10**

I leave you with the environmentalist Edward Abbey's excellent advice for those who are trying to change the world: "Do not burn yourself out. Be as I am–a reluctant enthusiast, a part-time crusader, a half-hearted fanatic. Save the other half of yourselves and your lives for pleasure."

You can't accomplish your goals unless you yourself are whole. Care for yourself and those around you. Take the time to do those things that replenish your energy and fill your heart with joy. Your children and our posterity will thank you for it.

The future is where we will spend the rest of our lives.

— **ANONYMOUS**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

For Melissa

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

. . . I'd always believed that a life of quality, enjoyment, and wisdom were my human birthright and would be automatically bestowed upon me as time passed. I never suspected that I would have to learn how to live—that there were specific disciplines and ways of seeing the world I had to master before I could awaken to a simple, happy, uncomplicated life. — **DAN MILLMAN**

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

EPILOGUE

SOME PARTING THOUGHTS

A lie can travel halfway around the world while the truth is putting on its shoes. — **MARK TWAIN**

Soon after I completed the first edition of *Turning Numbers into Knowledge* in 2001 I investigated a high profile example of bad statistics run amok. In this case, the tools and techniques described in this book, if widely applied, would have allowed businesses and investors to recognize these erroneous numbers for what they were before any harm was done.

DUBIOUS CLAIMS

Electricity used by computers was in the news awhile back, and some dubious claims were making the rounds:

- Claim # 1: The Internet accounted for eight percent of all electricity used in the United States around the year 2000.
- Claim # 2: Computers and information technology equipment, including the Internet, used 13 percent of all U.S. electricity in 2000—and this total was expected to grow to half of all electricity use ten to twenty years later.
- Claim #3: The "behind-the-wall" networking, server, and switching equipment needed to support a wireless personal digital assistant used as much electricity as a residential refrigerator.

Most major U.S. newspapers and business magazines, many respected institutions, and politicians of both political parties cited these assertions (the first one even came up in a Doonesbury comic strip at about the same time). Unfortunately, these claims are all incorrect.

These misconceptions originated in an article for *Forbes* magazine in May 1999 ("Dig More Coal—The PCs Are Coming"). They were perpetuated in a supporting report for the Greening Earth Society, in a *Wall Street Journal* op-ed piece, in congressional testimony, in articles for popular magazines, and in numerous interviews.

The claim that computers use lots of electricity spread quickly, driven by a superficially plausible story line and an ostensibly related high-profile crisis in the California electricity sector. *Forbes,* a respected magazine, lent credibility to the argument, as did the writer and publisher George Gilder, who published the *Digital Power Report* (which gave stock recommendations for investors assuming that these assertions were true) about one year after the *Forbes* article came out.

The trade press, the investment community, and the popular media found the key claims in the *Forbes* article irresistible. They repeated those claims frequently, often without citing the original source—in effect, enshrining them as conventional wisdom. Leaders in business, government, and academia were misled by this barrage of media attention and cited the statistics widely, ensuring their continued proliferation (Koomey et al. 2002).

For example, reports of six major investment banks published from May to September 2000 accepted the assertions in the *Forbes* article as fact (see **Table EP.1**). Each report claimed to present unique insights on this topic, but each relied heavily on the *Forbes* arguments for justification (sometimes attributing their stock picks to that source, sometimes not).

THE REAL STORY

Over several years, my colleagues and I demonstrated in peer-reviewed journal articles, congressional testimony, conference papers, and technical reports that

TABLE EP.1. Investment banking reports relying on the *Forbes* information technology electricity use figures

Publication and date	Quotation
(1) Deutsche Bank May 2000	"Mark Mills estimates that by 1999, the growth in (sic) Internet and related IT equipment now consumes 13% of our electricity supplies."
(2) Banc of America Securities June 2000	"Internet-related demand for power represented 8% to 13% of electricity consumption in 1999...It is estimated that by 2010, one-half of U.S. electric consumption will be related to the Internet in some way."
(3) Stephens, Inc. August 2000	"The percentage of electricity consumed directly by the Internet is currently estimated to be 10% in the U .S., up from roughly zero in 1993, and there are no signs of slowing growth in the pervasiveness of the web. Some estimates project that the Internet and the equipment to support its growth will consume 50% of domestic power within 10 years."
(4) Credit Suisse First Boston Corporation August 14, 2000	"This increase in dependence on electricity is attributable to the ongoing transformation of our economy from an industrial basis to one centered on knowledge, intellectual capital and technology. According to one study, nearly 8% of all the electricity generated in the United States is consumed by PCs, the Internet, and its related infrastructure. By 2020, that same study projects this consumption to grow to nearly 50%."
(5) JP Morgan Sept 14, 2000	"... information technology (IT) and telecom should account for an increasingly large piece of the total energy pie (up from about 16% today)."
(6) Salomon Smith Barney Sept. 25, 2000	"In 1995, the U.S. Department of Energy estimated that personal computers consumed approximately 3% of U.S. electricity supplies. Mark Mills, a well-known technology consultant, estimates that in 1999 Internet and related IT equipment consumed 13% of our electricity supplies."

(1) Tirello Jr., Edward J., Barbara Coletti, and Christopher R. Ellinghaus. 2000. *Convergence Redefined: The Digital Economy and the Coming Electricity Capacity Emergency.* Deutsche Banc Alex. Brown. May 12.

(2) Anderson, Hugh M. M., Russell L. Leavitt, James P. LoGerfo, and Daniel L. Tulis. 2000. *The Power of Growth.* NY, NY: Banc of America Securities. June.

(3) Stephens Inc. 2000. *Emerging Power Technology: Industry Report.* Little Rock, Arkansas: Stephens, Inc. August 11.

(4) Pencak, Marko, Neil Stein, Cameron Jeffreys, Christopher E. Vroom, Nat Schindler, Andrew Levi, Steven Parla, and Paul Patterson. 2000. *Energy Technology: An Overview.* Ontario, Canada: Credit Suisse First Boston Corporation. August 14.

(5) Feygin, Anatol, Waqar Syed, and Kyle Rudden. 2000. I*ndustry Analysis: We need more juice—IT and Telecom Growth Fuel Energy Demand—E&P, Natural Gas Pipeline, and Generators Should Benefit.* New York, NY: JP Morgan. September 14.

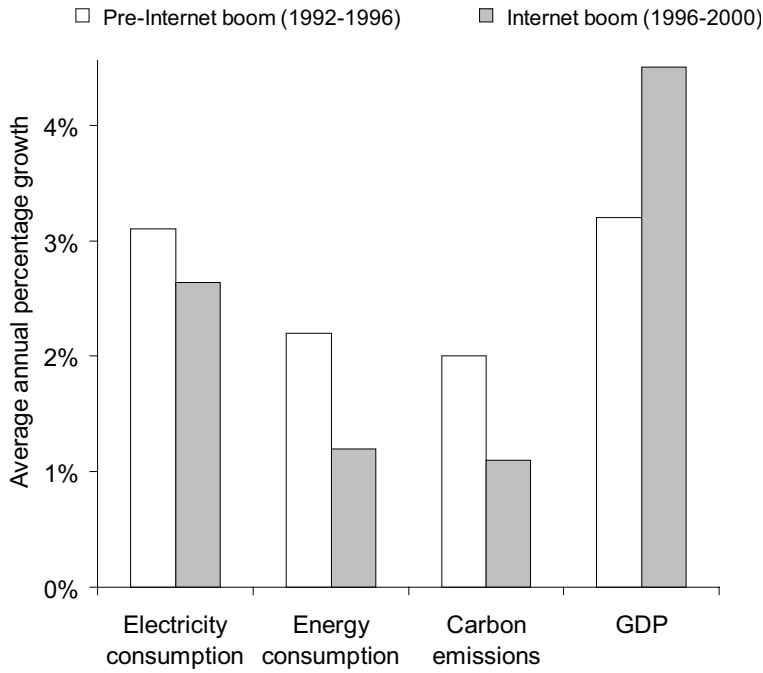
(6) Niles, Raymond C., Joanne M. Fairechio, and Christopher R. Ellinghaus. 2000. *The Power Curve.* New York, NY: Salomon Smith Barney. September 25.

the figures from the *Forbes* article were incorrect (see the citations in the Further Reading section for this Epilogue). We investigated these claims and in virtually every case they represented overestimates of electricity used by office equipment. For example, the Greening Earth Society report (Mills 1999) stated the assumption that a personal computer (PC) used 1,000 watts. Measurements I took at the time indicated that PCs typically used 60–80 watts. Commonly used 17-inch cathode ray tube monitors used about 90 watts in active mode, but the flat panel displays of comparable size (which were becoming more widely used around 2000) used a half or a third that amount. After accounting for other peripherals and "behind the wall" equipment, we concluded that 200 watts (not 1,000 watts) was a reasonable estimate for the active power of PCs, a factor-of-five reduction from the estimates in the *Forbes* article.

One would expect to find small differences between any such estimates, but when errors are all large and in the same direction it is likely that there is a bias in the analysis. Our studies found that the *Forbes* article overestimated power used by the Internet by at least eight times (i.e., the Internet used less than 1% of US electricity in 2000). It also overstated the total power used by office equipment by about four times—the actual percentage of US electricity used by all office equipment was 3% in 2000 (for details, see the Web site at <http://enduse.lbl.gov/projects/infotech.html>). The claim about wireless Palm Pilot electricity use was too high by a factor of two thousand (Koomey et al. 2004). The consulting firms Arthur D. Little and RAND published other independent analyses, funded by the U.S. Department of Energy, that confirmed our findings (Baer et al. 2002, Roth et al. 2002).

18 We also checked independent data sources, which showed that something was amiss with these claims. Joe Romm of the Center for Energy and Climate Solutions plotted **Figure EP.1** from Energy Information Administration data. The figure shows annual growth rates for U.S. electricity use, primary energy use, gross domestic product (GDP), and carbon emissions for the 1992 to 1996 and 1996 to 2000 periods. While GDP grew faster in the second period, electricity use, energy use, and CO_2 emissions all grew more slowly in that period. If the thesis of the *Forbes* article was correct, electricity demand growth rates at the end of the 1990s (the heyday of the Internet) would have

FIGURE EP.1 Comparison of average annual percentage growth rates in electricity use, energy use, carbon emissions, and Gross Domestic Product (GDP) for the pre-Internet and Internet boom periods.



SOURCE: Joe Romm of the Center for Energy and Climate Solutions, based on Energy Information Administration data, as cited in Koomey et al. (2002).

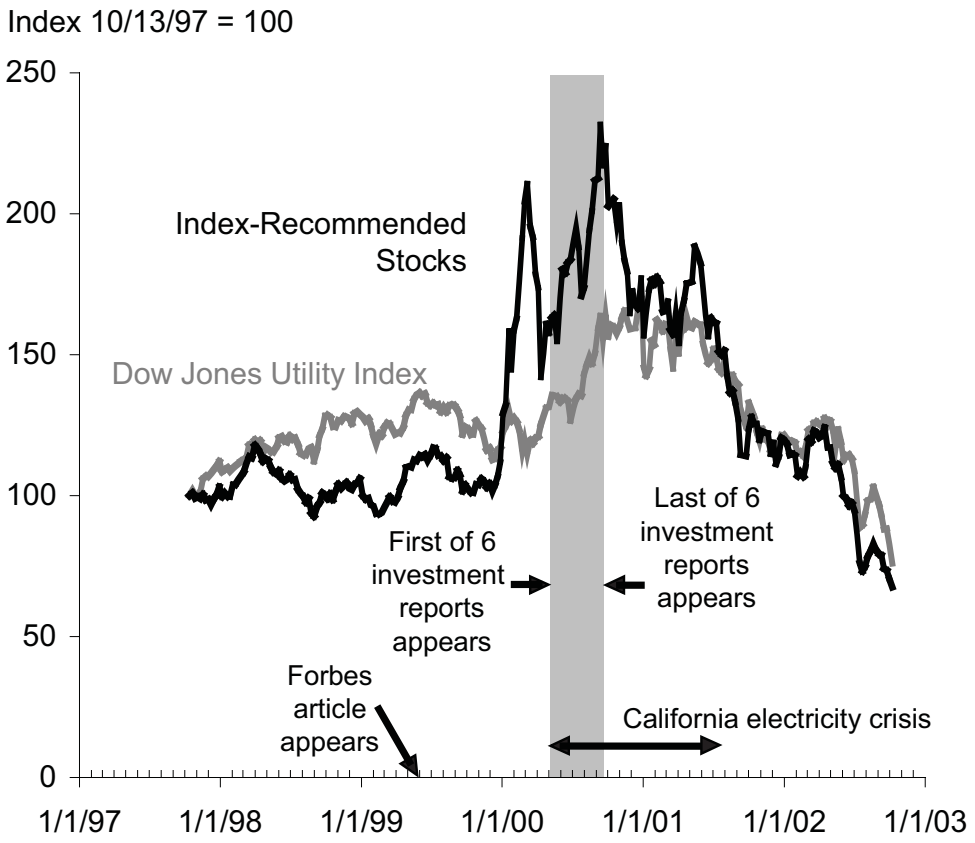
increased, but in fact the opposite occurred. These data contradict the assertion that demand growth was stronger with the widespread use of the Internet.

THE COST TO INVESTORS

In the absence of a formal "event study,"[85] no one knows for sure if these erroneous data and the associated investment recommendations cost investors money. I present some relevant data below so that readers can make their own judgments.

The stocks favored by the investment banking analysts (Table EP.1) were all in the area of electric power. Some were utilities and large energy trading com-

FIGURE EP.2 Price-weighted index of investment bank recommended stocks compared to the Dow Jones Utility Index from 10/13/1997 to 10/6/2002, indexed to 10/13/1997 = 100.



NOTE: Recommended stocks taken from reports listed in Table 1. Not all stocks were recommended by all reports. Stock prices downloaded from Yahoo! Finance in October 2002.

The ticker symbols for the complete list of recommended stocks from all the reports are: AES AMSC APC APCC APWR BLDP CAMZ CDTS CEG CPN CPST D DUK DVN DYN EIX ELON ENE ENER EPG ETR FCEL IMGC ITRI NI NRZ PCG PEG PLUG REI SATC SO SPIR WMB.

panies, others were smaller firms focusing on power quality, reliability, distributed generation, and power industry services. The looming electric reliability issues in California seemed to validate the focus on these firms.

Figure EP.2 plots an index of the 34 stocks recommended by the reports produced by the investment banks listed in Table 1. This index is weighted by stock price so that it is comparable to the Dow Jones Utility (DJU) index, also plotted here. The chosen stocks consistently underperformed the utility index until

about January 2000, when over a three month period their prices doubled. The price index of the recommended stocks then dropped by about one third, but rose again to make up those losses and more by the end of September 2000. It is suggestive (but not conclusive) that a large increase in the price of these stocks occurred over the period in which the aforementioned investment reports were published. Of course, the biggest troubles in California's restructured electricity markets also began to manifest themselves over the same period, but the hype generated by those events fed into the hype created by the investment reports, so they were self reinforcing.

Let's assume an investor had followed the recommendations in these six reports and bought 100 shares of each of the 34 touted companies back in September 2000 (when the last of these reports came out). By September 2002 he would have lost two-thirds of his investment. For comparison, the DJU index fell by 44%, and while few stock pickers did well during this period, most would regard losing 22 percentage points more than the DJU index over a two-year period as truly dismal investment performance.

THE ADVENTURE CONTINUES

Unfortunately, being refuted in the peer-reviewed literature didn't stop the authors of the *Forbes* article from continuing their crusade. They repeated their claims in a 2005 book (Huber and Mills 2005), without acknowledging their errors. This is another feature of some peddlers of erroneous statistics: they won't fess up to their mistakes. In addition, the quality controls on books are often far less stringent than those for peer-reviewed journals. And as this example shows, it takes many paragraphs (and a lot of work) to refute even just an errant sentence, so debunking of incorrect assertions always lags such claims.

This story is a familiar one. Another example is that of the book *The Skeptical Environmentalist,* which has been roundly criticized by virtually everyone who knows anything about the topics it covers, but that hasn't stopped its author from continuing to repeat his erroneous and misleading statements. By not admitting to and correcting their errors, these authors have violated the accepted code of conduct in science, best summarized by John Holdren, Past President of the American Association for the Advancement of Science, in his rebuttal to *The Skeptical Environmentalist:*

> The practice of science, which includes the packaging of findings from science for use in the public-policy arena, is governed by an unwritten code of conduct that includes such elements as mastering the relevant fundamental concepts before venturing into print in the professional or public arena, learning and observing proper practices for presenting ranges of respectable opinion and uncertainty, avoiding the selection of data to fit pre-conceived conclusions, reading the references one cites and representing their content accurately and fairly, and acknowledging and correcting the errors that have crept into one's work (some of which are, of course, inevitable) after they are discovered by oneself or by others.
>
> Most scientists follow this code of conduct as best they can out of self-respect and respect for the integrity of science itself. For those for whom these considerations might not be quite enough, there is little that can enforce the code other than concern with the cumulative harm to one's reputation and standing that comes from one's colleagues' awareness of a pattern of infractions, or fear of the public denunciation by colleagues that may follow in the rarer instances of someone's descending into more massive and willful disregard of accepted standards.

Violations of this code of conduct are anathema to scientists. Science is rare among human endeavors in that real progress over decades is easy to show. Better science is more accurate and powerful in its description of the natural world, as demonstrated by experiment, data collection, and the embodiment of scientific knowledge in technology. When serial obfuscators throw sand in the eyes of the scientific and policy process, it is an affront to all those who value truth over self-interest and muddy thinking.

INTEGRITY AND DECISION MAKING

I saw Scott McNealy, Chairman of Sun Microsystems, address the World Business Forum in October 2007. He returned again and again to the importance of integrity to success in business, because it enhances your ability to achieve measurable results, improves your credibility, and builds trust between you, your colleagues, and your competitors.

Webster's Ninth New Collegiate dictionary defines integrity as "firm adherence to a code of especially moral or artistic values: incorruptibility". Wikipedia (ac-

cessed January 31, 2008) defines it as "basing . . . one's actions on an internally consistent set of principles" <http://en.wikipedia.org/wiki/Integrity>. Analysts have integrity when they follow the code of conduct outlined by Holdren and the analytical principles outlined in this book. At the core of these principles is a respect for truth above all else, even when the truth contradicts long-held beliefs. **4**

Robert Ringer, one of my favorite authors, summarized his theory of reality as follows:

> This theory emphasizes, first of all, that reality isn't the way you wish things to be, nor the way they appear to be, but the way they actually are. Secondly, the theory states that you either acknowledge reality and use it to your benefit or it will automatically work against you.

Ringer uses the term "reality", but he could just as well have used the words "the truth" in his theory. He captures an essential lesson for those using analysis to improve decision making, and that is, after all, the name of the game.

Getting the numbers right really does matter. All business and policy decisions today are based at least in part on quantitative data, and no good can come of incorrect information being widely accepted. Annoyance, inconvenience, financial hardship, or even disaster can follow from important decisions based on faulty data.

We can't undo what's been done, but there is plenty that all of us can do from now on to ensure that we aren't being misled by bogus information. The widely cited but misleading statistics I mention here are just the tip of the iceberg. Your best defense against being fooled is to think for yourself and do your homework. Never make important decisions based on "common knowledge" unless you can verify its accuracy independently. Make sure you "follow the money" to determine whether someone has a vested interest in a particular outcome. And be skeptical about recommendations in investment reports, unless you know and trust the analyst's integrity and judgment. **11** **14** **24** **16**

You can improve your own analytical integrity by applying the techniques described in this book. Doing so can be time consuming, but if a set of facts is central to an important decision—for example, a business plan, a marketing strategy, or a new direction in research—then you need to put those principles to work.

CONCLUSIONS

People with integrity are the ones who succeed in the long run. The question remains, however: how can you best develop analytical and personal integrity? I offer these final thoughts:

- *Say what you mean and mean what you say.* If you promise something, always follow through—the recipient will reward your diligence when *you* need help next time.
- *Stick close to the data.* Don't make claims the data don't support, and don't let others do so. Never exaggerate, distort, or dissemble, or the results will come back to bite you in the butt (which is a more colorful way of rephrasing Ringer's "Theory of Reality").
- *Assume that others will replicate your work.* If it's interesting and important enough they surely will, so document it well and be accurate in the claims you infer from it.
- *Focus on completing the mission.* We've all encountered people whose personal issues, biases, and interests win out over the larger goals, and those are people to avoid if you can. True professionals can properly handle personal interests in a way that still leads to the best outcome for the team as a whole, but such professionals are the exception rather than the rule. You should be such an exception.

Analysis is, at its heart, a tool for building a better world. I wish you great success in putting it to work!

Always do right. This will gratify some people and astonish the rest.

— MARK TWAIN

FURTHER READING

NUMBERS TO KNOWLEDGE "TOP ELEVEN" LIST

Choosing my eleven favorites from the "Further Reading" section was not an easy task (I couldn't keep myself to just ten). I know you won't be disappointed with any of these:

1. Huff, Darrell. 1993. *How to Lie with Statistics*. New York, NY: W. W. Norton & Co., Inc. If this book doesn't make you a skeptical observer of numbers and statistics, nothing will. First published in the 1950s, it's still fresh and useful today.
2. Booth, Wayne C., Gregory G. Colomb and Joseph M. Williams. 1995. *The Craft of Research*. Chicago, IL: The University of Chicago Press. This superb book presents a framework for dissecting arguments that can be applied in most situations.
3. Ueland, Brenda. 2007. *If You Want to Write: A Book About Art, Independence, and Spirit*. St. Paul, Minnesota: Graywolf Press. There is no better guide to sparking your creative process.
4. Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. 2d Peterborough, Ontario: Broadview Press. This clearly written and accessible book is a marvelous introduction to logical argument.
5. U.S. Bureau of the Census. 2008. *The Statistical Abstract of the United States 2008*. 127th edition Washington D.C.: U.S. Government Printing Office <http://www.census.gov/compendia/statab>. One of the most important data sources for any U.S. researcher. The plot's a bit thin, but the content's superb.
6. Norman, Donald A. 1990. *The Design of Everyday Things*. New York, NY: Doubleday/Currency. This fascinating book explores how humans interact with technology, and gives guidance to designers on how to make their products a joy to use.
7. Schwartz, Peter. 1996. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York, NY: Doubleday. Lively and realistic scenarios are crucial tools for decision makers, and Schwartz describes how to create them in this classic book.
8. Heath, Chip, and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. New York, NY: Random House. This book contains insightful and practical guidance for compelling story telling.
9. Tufte, Edward R. *Envisioning Information*. (1990). *Visual Explanations: Images and Quantities, Evidence and Narrative*. (1997). *The Visual Display of Quantitative Information*, 2nd ed. (2001). Cheshire, CT: Graphics Press. OK, I'm cheating a bit by calling these one book, but they are so good you'll want them all <http://www.edwardtufte.com>.
10. Few, Stephen. 2004. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press. This book complements Tufte's work, but is more business-focused and practical.
11. Ringer, Robert J. 1974. *Winning Through Intimidation*. New York: Funk & Wagnalls. (Republished in 2004 as "*To Be or not to Be Intimidated? That is the Question*".) This book's philosophical and practical advice (and its humor) makes it one of my all-time favorites.

INTRODUCTION • THE INFORMATION EXPLOSION

Burke, James. 2007. *Connection*. New York, NY: Simon & Schuster. An outstanding book that discusses the history of technology in a brilliant and accessible way. Among other things, Burke explores the key role of the printing press in fostering later technological and scientific developments.

Cohen, Patricia Cline. 1985. *A Calculating People: The Spread of Numeracy in Early America*. Chicago, IL: The University of Chicago Press. This marvelous history describes the commercial, scientific, and philosophical origins of European and American fascination with numbers and statistics.

de Sola Poole, Ithiel. 1983. *Technologies of Freedom*. Cambridge, MA: The Belknap Press of Harvard University Press. This book made a first attempt to quantify the accelerating pace of information generation in modern society.

Koomey, Jonathan, Marshall van Alstyne, and Erik Brynjolfsson. 2007. "You've got spam." *The Wall Street Journal*. New York, NY. September 6. p. A16. This op-ed cites the statistic that 90% of 2007 email is spam and proposes an approach for solving the spam problem.

Shenk, David. 1997. *Data Smog: Surviving the Information Glut*. New York, NY: HarperCollins Publishers, Inc. This book documents the symptoms of information overload. It's worth a read just for the anecdotes, which are delightful.

PART I • THINGS TO KNOW

1 • Beginner's Mind

Heath, Chip, and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. New York, NY: Random House. This wonderful book describes "the curse of expertise" and its implications for how we interpret what's going on in the world.

Hyams, Joe. 1982. *Zen in the Martial Arts*. New York, NY: Bantam Books. The story of Joe Hyams' explorations into different martial arts is one of the inspirations for the book you are now reading. It's broken up into short sections, each with an important lesson he learned over the years. A quick read, and a great one.

2 • Don't Be Intimidated

Ringer, Robert J. 1974. *Winning Through Intimidation*. New York: Funk & Wagnalls. This book is one of my favorites. It's entertaining, interesting, and timeless. It's also a quick read, and is available in most big bookstores (it still sells well). It was republished in 2004 as *To Be or Not to Be Intimidated: That is the Question.*

Ueland, Brenda. 2007. *If You Want to Write: A Book About Art, Independence, and Spirit*. St. Paul, Minnesota: Graywolf Press. One of the classic books on creativity. Ueland writes about writing, but her advice applies equally well to most forms of human creation. It's beautifully written and easy to read.

3 • Information, Intention, and Action

Alonso, William, and Paul Starr, ed. 1987. *The Politics of Numbers*. New York, NY: Russell Sage Foundation. Alonso and Starr edit a collection of essays that brilliantly explores the political and practical implications of increasing reliance on fact-based decision making in the U.S. in the Twentieth Century.

Brooks Jr., Frederick P. 1982. *The Mythical Man-Month: Essays on Software Engineering*. Reading, MA: Addison-Wesley Publishing Company. This book gives sage and practical advice on programming that applies more generally to implementing ideas in the real world. A classic!

Gladwell, Malcolm. 2005. *Blink: The Power of Thinking Without Thinking*. New York, NY: Little, Brown and Company. This book shows how people use "gut feelings" to make important decisions (and why some are better at it than others).

Hammond, John S., Ralph L. Keeney, and Howard Raiffa. 1999. *Smart Choices: A Practical Guide to Making Better Decisions*. Boston, MA: Harvard Business School Press. A well-written and methodical guide to improving your decision making processes.

Hess, Peter, and Julie Siciliano. 1996. *Management: Responsibility for Performance*. New York, NY: Mc Graw-Hill, Inc. This book does a fine job of describing how to set goals, make decisions, and implement decisions in the business world. It also explicitly discusses the role of intuition, experience, and values in decision making.

Kawasaki, Guy. 1995. *How to Drive Your Competition Crazy: Creating Disruption for Fun and Profit*. New York, NY: Hyperion. Kawasaki harkens back to his time as chief evangelist for the upstart Apple Computer in the late 1970s and early 1980s, and gives some lessons for marketing products in the face of competition from huge and well-funded rivals. The book is loaded with great anecdotes.

Kim, Jane J. 2005. "Stock Research Gets More Reliable." *The Wall Street Journal*. New York, NY. June. p. D1. Changes in the way research departments are structured in the big investment firms have yielded some improvements in their investment research reports.

Laderman, Jeffrey M. 1998. "Wall Street's Spin Game: Stock Analysts Often Have a Hidden Agenda." In *Business Week*. October 5. pp. 148-156.

Morgan, M. Granger, and Max Henrion. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY: Cambridge University Press. A "must read" book for anyone trying to understand how uncertainty affects decision making.

Norman, Donald A. 2002. *The Design of Everyday Things*. New York, NY: Basic Books. This fascinating book explores how humans interact with technology, and gives guidance to designers on how to make their products a joy to use.

Schwartz, Peter. 1996. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York, NY: Doubleday. Lively and realistic scenarios are crucial tools for decision makers, and Schwartz describes how to create them in this classic book.

Story, Louise. 2007. "How Many Site Hits? Depends Who's Counting." *The New York Times (online)*. New York, NY. October 22. Measurement of site traffic is still difficult, even after more than a decade of improvements in the World Wide Web.

4 • Peer Review and Scientific Discovery

Borchert, Donald M., ed. 2006. *The Encyclopedia of Philosophy,* 2nd ed. New York, NY: Macmillan Reference Books. Volumes 1 through 10. This encyclopedia is the standard reference source for anyone investigating philosophical concepts. It contains articles by leading scholars on everything from logic to epistimology, energy to truth, liberty to idealism, ethics to virtue, and almost anything else you can think of.

Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. 2d Peterborough, Ontario: Broadview Press. This clearly written and accessible book is a wonderful introduction to logical argument. Hughes uses myriad examples to illustrate the logical fallacies that

bedevil most arguments. He also clearly describes the difference between inductive and deductive logic.

Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago: University of Chicago Press. This book redefined contemporary views of how science progresses, and is a landmark in the history of science.

Murray, David, Joel Schwartz, and S. Robert Lichter. 2001. *It Ain't Necessarily So: How Media Make and Unmake the Scientific Picture of Reality*. Lanham, MD: Rowman & Littlefield Publishers, Inc. This book offers a nuanced and informative look at how the media reports on scientific research.

Pirsig, Robert M. 2006. *Zen and the Art of Motorcycle Maintenance*. NY: HarperTorch. This book is a popular excursion into the relationship between science and art. While not always philosophically rigorous, the book is still entertaining, thoughtful, and provocative.

Sokal, Alan D. 1997. "A Plea for Reason, Evidence, and Logic." *New Politics*. vol. 6, no. 2. p. 126. This short transcript of a talk by Sokal is a brilliantly clear and concise exposition of the scientific worldview, in the context of his now famous parody of "post-modernist" relativism. Sokal created a controversy in 1996 by submitting an article that he knew was total nonsense to a journal called *Social Text*, and the journal published it. Sokal suspected correctly that by liberally sprinkling the article with phrases that flattered the editors' predispositions, he would make publication likely. He then exposed the hoax. For the complete set of articles and related debates, see <http://www.physics.nyu.edu/faculty/sokal>.

Thomson. 2008. Science Citation Index, online version: <http://scientific.thomson.com/products/scie>. This web site is the on-line version of the science citation index.

Wilber, Ken. 1998. *The Marriage of Sense and Soul: Integrating Science and Religion*. New York, NY: Broadway Books. Wilber is a science-minded philosopher who has real insight into how scientific principles can help humans to more fully explore all aspects of reality.

PART II • BE PREPARED

5 • Explore Your Ideology

Flew, Anton, ed. 1984. *Dictionary of Philosophy*. London: Pan Books, Ltd.

6 • Get Organized

Covey, Stephen R. 2004. *The 7 Habits of Highly Effective People*. 15th Anniversary edition. NY, NY: Free Press. Covey got me to think systematically about how I organize my time and my work.

Culp, Stephanie. 1990. *Conquering the Paper Pile-up: How to Sort, Organize, File, and Store Every Piece of Paper in Your Home and Office*. Cincinnati, Ohio: Writer's Digest Books. Practical advice for getting organized.

Lively, Lynn. 1996. *Managing Information Overload*. The WorkSmart Series. New York, NY: AMACOM/American Management Association. This book is brief but important. Its advice on controlling information flows is especially pertinent in the information age.

Morgenstern, Julie. 2004. *Organizing from the Inside Out: The Foolproof System for Organizing Your Home, Your Office and Your Life*. 2nd ed. New York, NY: Holtzbrinck Publishers (now MacMillan). This is one of the best books on organizing that I've found. It's a top seller.

7 • Establish a Filing System

Culp, Stephanie. 1990. *Conquering the Paper Pile-up: How to Sort, Organize, File, and Store Every Piece of Paper in Your Home and Office*. Cincinnati, Ohio: Writer's Digest Books. A helpful book focusing specifically on dealing with paper clutter.

Lively, Lynn. 1996. *Managing Information Overload*. The WorkSmart Series. New York, NY: AMACOM/American Management Association. Advice on remaining organized in the face of a flood of information.

Morgenstern, Julie. 2004. *Organizing from the Inside Out: The Foolproof System for Organizing Your Home, Your Office and Your Life*. 2nd ed. New York, NY: Holtzbrinck Publishers (now MacMillan). Has a fine chapter on "traditional offices and filing systems."

Schwartz, Peter. 1996. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York, NY: Doubleday. See the section on "Information hunting and gathering".

8 • Build a Toolbox

The Economist. 2003. *Numbers Guide: The Essentials of Business Numeracy*. Princeton, NY: Bloomberg Press. This terrific book teaches numerical tools and techniques for business applications.

Gonick, Larry, and Woollcott Smith. 1993. *The Cartoon Guide to Statistics*. New York, NY: Harper Perennial, a division of Harper Collins Publishers. This book is a fun way to learn basic statistical concepts.

Harte, John. 1985. *Consider a Spherical Cow: A Course in Environmental Problem Solving*. Los Altos, CA: William Kaufmann, Inc. This beautiful book contains tools relevant to solving problems in environmental science. The sequel to it (*Consider a Cylindrical Cow: Further Adventures in Environmental Problem Solving*) was published in 2001 by University Science Books, in Sausalito, California.

Huff, Darrell. 1993. *How to Lie with Statistics*. New York, NY: W. W. Norton & Co., Inc. Reread this classic book every few years, to keep your toolbox fresh.

Huff, Darrell. 1999. *The Complete How to Figure It*. New York, NY: W. W. Norton & Co., Inc. An excellent sourcebook, with hundreds of examples of how to solve different problems in the real world.

Jones, Gerald Everett. 2006. *How to Lie with Charts*. 2nd ed. Charleston, SC: BookSurge Publishing. A fine source for ideas on graphical and tabular presentation.

Kernighan, Brian W., and Rob Pike. 1999. *The Practice of Programming*. Reading, MA: Addison-Wesley. This marvelous book on the art of computer programming is somewhat technical, but is exceptional for its advice on programming style and documentation. It also offers many excellent suggestions for debugging and testing calculations.

Swartz, Clifford E. 1993. *Used Math: for the First Two Years of College Science*. 2nd ed. College Park, MD: American Association of Physics Teachers. Summarizes mathematics that can be useful for more advanced analysts.

Tal, Joseph. 2001. *Reading Between the Numbers: Statistical Thinking in Everyday Life*. New York, NY: McGraw-Hill. This book offers both conceptual understanding of key statistical concepts and intensely practical guidance for real-world use of those concepts.

9 • Put Facts at Your Fingertips

Annual Reviews. 4139 El Camino Way, PO Box 10139. Palo Alto, CA 94303-0139. 650/493-4400 <http://www.annualreviews.org>.

Berkman, Robert I. 2000. *Find It Fast: How to Uncover Expert Information In Any Subject*. 5th ed. New York, NY: Harper & Rowe, Publishers. A helpful guide to digging out data from the myriad disparate sources out in the world.

Blocksma, Mary. 1989. *Reading the Numbers: A Survival Guide to the Measurements, Numbers, and Sizes Encountered in Everyday Life*. New York, NY: Penguin USA. This book is somewhat hard to find, but it explains how to interpret and use a variety of data that we encounter in everyday life.

Borchert, Donald M., ed. 2006. *The Encyclopedia of Philosophy*. 2nd ed. New York, NY: Macmillan Reference Books. Volumes 1 through 10. The best place to go when you need some background on an intellectual concept.

Joyce, C. Alan, ed. 2007. *The World Almanac and Book of Facts 2008*. Mahwah, New Jersey: World Almanac Books. A "must have" reference. It comes out every year, and is chock full of useful information <http://www.worldalmanac.com>.

Lide, David R., ed. 2007. *CRC Handbook of Chemistry and Physics*. 88th ed. Boca Raton, Florida: CRC Press, Inc. The best source for physical constants and conversions. CRC also has many other technical and scientific reference works. Earlier editions (still useful) can be bought used.

Simpson, John and Edmund Weiner, eds. 2002. *The Oxford English Dictionary*. Oxford, UK: Oxford University Press. 2d edition. This reference work takes up 20 volumes and 22,000 pages! This version costs about $1,000 used, but you can get a complete version with *very* small type for about $40 used (it comes with its own powerful magnifying glass). A better choice is the CD-ROM version, available for both Windows and Macintosh for about $200.

Sutcliffe, Andrea. 1996. *Numbers: How many, How Long, How Far, How Much*. New York, NY: Stonesong Press, Inc. Another good sourcebook for all sorts of helpful data.

U.S. Bureau of the Census. 2008. *The Statistical Abstract of the United States 2008*. 127th edition Washington D.C.: U.S. Government Printing Office. One of the most important data sources for any U.S. researcher.

Worldwatch Institute. 2007. *Vital Signs 2007–2008: The Trends That Are Shaping Our Future*. New York, NY: W.W. Norton and Co. Tons of good time-series data on environmental and technology trends. Worldwatch also sells these data in electronic form in both Windows and Macintosh formats. <http://www.worldwatch.org>

10 • Value Your Time

Covey, Stephen R. 2004. *The 7 Habits of Highly Effective People*. 15th Anniversary edition. NY, NY: Free Press. Covey is a highly successful "personal growth" guru who doles out advice for a living. His book helps you explore how your life patterns thwart or facilitate success.

DeMarco, Tom, and Timothy Lister. 1999. *Peopleware: Productive Projects and Teams*. 2nd ed. New York, NY: Dorset House Publishing Company. A superb book that uses real-world experience to explode dozens of management myths. If you manage creative people, you owe it to them and to yourself to read this book.

Dominguez, Joe and Vicki Robin. 1999. *Your Money or Your Life: Transforming Your Relationship with Money and Achieving Financial Independence*. New York, NY: Penguin Books. Includes a clear and persuasive statement of how money, time, and your life are intimately connected.

Hyams, Joe. 1982. *Zen in the Martial Arts*. New York, NY: Bantam Books. The chapter titled "Do not disturb" describes how Hyams and his martial arts colleagues take command of their time.

Thich Nhat Hanh. 1991. *Peace is Every Step: The Path of Mindfulness in Everyday Life*. New York, NY: Bantam Books. This book advocates awareness in every waking moment of every day, so you never forget how amazing the world is. Sadly, most of us forget this all the time.

Thoreau, Henry David. 1971. *Thoreau: Walden and Other Writings*. New York, NY: Bantam Books, Inc. Joseph Wood Krutch, ed. Contains Thoreau's musings during his stay at Walden Pond.

PART III • ASSESS THEIR ANALYSIS

General

Best, Joel. 2001. *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*. Berkeley, CA: University of California Press. This brilliant book gives insight into how to recognize mutant statistics, apples and oranges comparisons, and other common issues with numbers. It also teaches a critical approach to avoiding such pitfalls.

Huff, Darrell. 1993. *How to Lie with Statistics*. New York, NY: W. W. Norton & Co., Inc. There's no better exposé of commonly used tricks that mislead and misrepresent than this book.

Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. 2d Peterborough, Ontario: Broadview Press. Hughes presents examples and exercises to explore and categorize the logical fallacies that befall most arguments.

Lieber, Hugh Gray, and Lillian R. Lieber. 1944. *The Education of T. C. MITS*. New York: W. W. Norton & Company, Inc. This book is an entertaining excursion into mathematics, logic, and argument. Though it's dated in some ways, it's still worth reading.

Moroney, M. J. 1953. *Facts From Figures*. Harmondsworth, Middlesex, UK: Penguin Books Ltd. Another book from long ago, it's one of the best I've seen at explaining the practice of statistical analysis in the real world. It uses dozens of examples to illustrate statistical concepts, and while it can be technical at times, even lay readers will learn quickly from it.

11 • The Power of Critical Thinking

Gilovich, Thomas. 1991. *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. New York, NY: The Free Press. This book documents, categorizes, and explains errors of reasoning common to human beings. It also describes how best to challenge dubious beliefs once they become manifest.

Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. 2d Peterborough, Ontario: Broadview Press. On pages 108-110 of this outstanding book, Hughes gives his seven rules for assessing arguments.

Hammond, John S., Ralph L. Keeney, and Howard Raiffa. 1999. *Smart Choices: A Practical Guide to Making Better Decisions*. Boston, MA: Harvard Business School Press. A methodical guide to thinking through your choices.

Mayfield, Marlys. 2006. *Thinking for Yourself: Developing Critical Thinking Skills Through Reading and Writing*. 7th ed. Belmont, CA: Heinle Publishing. A superb textbook that teaches clear writing and critical thinking using examples and exercises.

Stebbing, L. Susan. 1961. *Thinking to Some Purpose*. Harmondsworth, Middlesex, UK: Pelican. This book marries critical thinking and rhetoric in an exemplary way.

12 • Numbers Aren't Everything

Bella, David A. 1979. "Technological Constraints on Technological Optimism." *Technological Forecasting and Social Change*. vol. 14, p. 15. An article that uses the methods of science and mathematics to show why social and environmental problems generated by technology can be difficult to solve.

Jonas, Hans. 1973. "Technology and Responsibility: Reflections on the New Task of Ethics." *Social Research*. vol. 40, no. 1. p. 31. This article addresses ethical issues that affect modern technological societies that for the first time have the power to effect changes in the natural environment in a global way. As technology changes what it means to be human, Jonas argues that contemporary ethics must keep pace.

13 • All Numbers Are Not Created Equal

DeCanio, Stephen. 2003. *Economic Models of Climate Change: A Critique*. Basingstoke, UK: Palgrave-MacMillan. A cogent and readable analysis of the problems with economic models.

Schwartz, Peter. 1996. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York, NY: Doubleday. The classic book on scenario analysis. A "must read" book for anyone trying to think systematically about the future.

14 • Question Authority

Brandt, Allan M. 2007. *The Cigarette Century: The Rise, Fall, and Deadly Persistence of the Product that Defined America*. New York, NY: Basic Books. A brilliant and comprehensive history of the anti-social behavior of one of America's most important industries.

Cialdini, Robert B. 2006. *Influence: The Psychology of Persuasion*. 1st Collins Business Essentials edition. New York, NY: Collins. An extraordinary book that documents how "compliance professionals" influence people's opinions and actions.

Crossen, Cynthia. 1994. *Tainted Truth: The Manipulation of Fact in America*. New York, NY: Simon & Schuster. Another eye opener that shows how people and institutions can and do play with numbers to convince and confuse.

Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. 2d Peterborough, Ontario: Broadview Press. Hughes presents detailed criteria for assessing the strength of appeals to authority.

Robbins, John. 1998. *Diet for a New America*. Tiburon, CA: HJ Kramer Inc. Whatever you think about Robbins' advocacy of a vegetarian diet, his description of how the dairy and meat industries got schools and other public institutions to distribute their marketing materials is a real eye opener.

16 • Don't Believe Everything You Read

Brandt, Allan M. 2007. *The Cigarette Century: The Rise, Fall, and Deadly Persistence of the Product that Defined America*. New York, NY: Basic Books. The stories in this book will make you skeptical of pronouncements by those with an axe to grind.

Crossen, Cynthia. 1994. *Tainted Truth: The Manipulation of Fact in America*. New York, NY: Simon & Schuster. Crossen summarizes how the results of research are often manipulated for self-serving ends.

Ewen, Stuart. 1996. *PR! A Social History of Spin*. New York, NY: Basic Books, a member of the Perseus Books Group. This book recounts the development of public relations and propaganda in the Twentieth Century. It is an intriguing and well-told tale.

Huff, Darrell. 1993. *How to Lie with Statistics*. New York, NY: W. W. Norton & Co., Inc. If this book doesn't make you a skeptical observer of numbers and statistics, nothing will. First published in the 1950s, it's still fresh and useful today.

17 • Go Back to the Questions

Crossen, Cynthia. 1994. *Tainted Truth: The Manipulation of Fact in America*. New York, NY: Simon & Schuster. Crossen demonstrates how research has become a tool for hucksters to make friends and influence people.

Weiss, Michael J. 1994. *Latitudes and Attitudes: An Atlas of American Tastes, Trends, Politics, and Passions*. New York, NY: Little Brown and Company. This book is a beautiful summary of consumer and market data. It shows the power of comprehensive consumer surveys to describe consumer preferences across the U.S.

18 • Reading Tables and Graphs

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press. Fourteenth printing. Reading this book will change the way you look at tables and graphs, and it is required reading for those learning this craft.

19 • Distinguish Facts from Values

Booth, Wayne C., Joseph M. Williams and Gregory G. Colomb. 2003. *The Craft of Research*. 2nd ed. Chicago, IL: The University of Chicago Press. This superb book presents a framework for dissecting arguments that can be applied in most situations. The authors were trained in rhetoric and literature, but their way of thinking about claims and evidence applies equally well to technical topics.

Fuchs, Victor R., Alan B. Krueger and James M. Poterba. 1997. *Why Do Economists Disagree About Policy? The Roles of Beliefs About Parameters and Values*. Princeton University, Industrial Relations Section. Working Paper #389. August.

Jonas, Hans. 1973. "Technology and Responsibility: Reflections on the New Task of Ethics." *Social Research*. vol. 40, no. 1. p. 31. This article, more than any other I read in graduate school, got me to think carefully about how technology can challenge some of our most deeply held values.

Marshall, Jonathan. 1997. "Economists Can't Seem to Agree on Anything." *The SF Chronicle*, August 25, 1997, p. B2.

Reich, Robert. 1997. *Locked in the Cabinet*. New York, NY: Alfred A. Knopf. Regardless of your political leanings, you'll enjoy Reich's tales of political intrigue, error, and innuendo in the Clinton White House.

Wilber, Ken. 1998. *The Marriage of Sense and Soul: Integrating Science and Religion*. New York, NY: Broadway Books. Wilber provides a philosophical framework for understanding how facts should relate to values, using the term "exterior" to describe the measurable world of facts and the word "interior" to describe the world of ideas and values.

20 • The Uncertainty Principle and the Mass Media

Cialdini, Robert B. 2006. *Influence: The Psychology of Persuasion*. 1st Collins Business Essentials edition. New York, NY: Collins. Cialdini's book shows how "compliance professionals" use their knowledge of the human psyche to influence people's decisions.

Crossen, Cynthia. 1994. *Tainted Truth: The Manipulation of Fact in America*. New York, NY:

Simon & Schuster. Crossen gives dozens of examples of how the results of research and reporting can be manipulated for good or ill.

Fuller, Jack. 1996. *News Values: Ideas for an Information Age*. Chicago, IL: University of Chicago Press. Fuller explores how new technologies continue to change the role of the journalist in modern society.

Levering, Robert, and Milton Moskowitz. 1994. *The 100 Best Companies to Work for in America*. Plume Publishing. This is the 1994 edition of their book, which was first published in 1984.

PART IV • CREATE YOUR ANALYSIS

21 • Reflect

Hyams, Joe. 1982. *Zen in the Martial Arts*. New York, NY: Bantam Books.

Mattimore, Bryan W. 1994. *99% Inspiration: Tips, Tales & Techniques for Liberating Your Business Creativity*. New York, NY: AMACOM (A Division of American Management Association).

Ueland, Brenda. 2007. *If You Want to Write: A Book About Art, Independence, and Spirit*. St. Paul, Minnesota: Graywolf Press.

22 • Get Unstuck

Elffers, Joost. 2002. *Play with Your Food*. New York, NY: Metro Books. This book encourages you to think about food in a way you never have before. If you really get stuck, get this book and play until you're ready to get back to the problem.

Mattimore, Bryan W. 1994. *99% Inspiration: Tips, Tales & Techniques for Liberating Your Business Creativity*. New York, NY: AMACOM (A Division of American Management Association). This book is a treasure trove of ideas for becoming more productive.

Polya, G. 2004. *How to Solve It: A New Aspect of Mathematical Method*. Princeton, N.J.: Princeton University Press. A serious but approachable book about problem solving. It's a real classic!

23 • Inquire

Berkman, Robert I. 2000. *Find It Fast: How to Uncover Expert Information In Any Subject, Online or in Print*. New York, NY: Collins. One researcher's tricks to digging out information in the real world.

Borchert, Donald M., ed. 2006. *The Encyclopedia of Philosophy*. 2nd ed. New York, NY: Macmillan Reference Books. Volumes 1 through 10. This set is an essential desk reference that describes the historical development of ideas. It can lend context to your research available nowhere else.

U.S. Bureau of the Census. 2008. *The Statistical Abstract of the United States 2008*. 127th edition Washington D.C.: U.S. Government Printing Office. One of the most important data sources for any U.S. researcher.

The Encyclopedia Brittanica. The hard copy of this encyclopedia costs about $1400, but the electronic versions are a much better deal. A multimedia CD-ROM version costs less than $100 (for both Windows and Macintosh), and a subscription-based web version <http://www.eb.com> is also available. Other encyclopedias offer similar deals.

24 • Be a Detective

Sir Arthur Conan Doyle. *A Study in Scarlet*. This novel, in which Conan Doyle created the modern detective, first brought Sherlock Holmes into English literature. It's a gripping, well-told story, and one in which the scientific method features prominently. Contained in Baring-Gould, William S. ed. 1976. *The Annotated Sherlock Holmes*. Volume I. New York, NY: Clarkson N. Potter, Inc./Publisher. 15th printing. pp. 143-234.

25 • Create Consistent Comparisons

Koomey, Jonathan G., Deborah E. Schechter and Deborah Gordon. 1993. *Cost-Effectiveness of Fuel Economy Improvements in 1992 Honda Civic Hatchbacks*. Energy and Resources Group, University of California, Berkeley. Presented at the 72nd Annual meeting of the National Research Council's Transportation Research Board, January 10-14, 1993, Washington, DC. May 1993. Download the report in PDF format at <http://enduse.lbl.gov/Info/Honda-abstract.html>.

26 • Tell a Good Story

Denning, Stephen. 2005. *The Leader's Guide to Storytelling: Mastering the Art and Discipline of Business Narrative*. San Francisco, CA: Jossey-Bass, A Wiley Imprint. Denning spent years studying how stories could help leaders change organizations, and this excellent book summarizes what he learned.

Ghanadan, Rebecca, and Jonathan Koomey. 2005. "Using Energy Scenarios to Explore Alternative Energy Pathways in California." *Energy Policy*. vol. 33, no. 9. June. pp. 1117-1142. This refereed journal article gives an extensive description of the scenario creation process, accompanied by graphics that help illustrate the approach.

Heath, Chip, and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. New York, NY: Random House. Simple, unexpected, concrete, credible and emotional stories are the ones that stick, and this book will help you create them.

Morgan, M. Granger, and Max Henrion. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY: Cambridge University Press. Analyzing uncertainty is key to technical story telling.

Schwartz, Peter. 1996. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York, NY: Doubleday. The classic book on building scenarios.

27 • Dig into the Numbers

Cook, Terry. 1995. "It's 10 O'Clock: Do You Know Where Your Data Are?" *Technology Review*. January. p. 48. An insightful article that discusses some unprecedented problems with data that we face in the information age.

28 • Make a Model

Bernstein, Peter L. 1998. *Against the Gods: The Remarkable Story of Risk*. New York, NY: John Wiley and Sons. This wonderful book tells how modern society learned to "put the future at the service of the present" by mastering risk and the tools needed to understand it.

Keepin, William (Bill). 1983. *A Critical Appraisal of the IIASA Energy Scenarios*. International Institute for Applied Systems Analysis. WP-83-104. October. The complete documentation for Keepin's criticisms of the IIASA scenarios.

Keepin, William and B. Wynne. 1984. "Technical analysis of IIASA energy scenarios." *Nature*. vol. 312, no. 5996. December 20. pp. 691–695. This is the journal article summary of Keepin's work at IIASA.

Morgan, M. Granger and Max Henrion. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY: Cambridge University Press. This book gives sage advice about appropriate uses for models in analysis, and is one of the few places where large complex models are taken to task for their failings.

Ringer, Robert J. 1974. *Winning Through Intimidation*. New York: Funk & Wagnalls. A funny and entertaining look at one man's transformation from loser to winner, in part through his clever use of models.

Starfield, Anthony M., Karl A. Smith and Andrew L. Bleloch. 1990. *How to Model It: Problem Solving for the Computer Age*. New York, NY: McGraw-Hill, Inc. This book is an excellent, albeit technical, guide to modeling practice in the real world, particularly computer modeling. It teaches problem solving skills essential for good modeling, and focuses mainly on scientific applications, though its approach is useful for non-scientists as well. It is unfortunately out of print, but many libraries will have it.

Varian, Hal R. 1997. *How to Build an Economic Model in Your Spare Time*. Berkeley, CA: University of California. June 11. (http://www.sims.berkeley.edu/~hal/Papers/how.pdf). An entertaining and readable guide to making economic models, with plenty of advice reflecting Varian's lifetime of real-world experience.

29 • Reuse Old Envelopes

Harte, John. 1985. *Consider a Spherical Cow: A Course in Environmental Problem Solving*. Los Altos, CA: William Kaufmann, Inc. The first short chapter in this classic book asks the question "How many cobblers are there in the U.S.?" and then proceeds to investigate ever more detailed puzzles in environmental science. This book, and the course in which it is used, were inspirational to me in graduate school. The sequel to it (*Consider a Cylindrical Cow: Further Adventures in Environmental Problem Solving*) was published in 2001 by University Science Books, in Sausalito, California.

Herendeen, Robert A. 1998. *Ecological Numeracy: Quantitative Analysis of Environmental Issues*. New York, NY: John Wiley & Sons, Inc. This book contains many nice examples of back-of-the envelope calculations, and gives a somewhat technical but still useful introduction to issues of energy and the environment.

Hershey, Robert L. 1982. *How to Think with Numbers*. Los Altos, CA: William Kaufmann, Inc. This book presents more than fifty specific problems to help you hone your calculational skills.

Hershey, Robert L. 1997. "Thinking Big." *The Washington Post*, September 10, 1997, p. H8. A short article talking about scientific notation and its usefulness for treating really BIG numbers.

Mattimore, Bryan W. 1994. *99% Inspiration: Tips, Tales & Techniques for Liberating Your Business Creativity*. New York, NY: AMACOM (A Division of American Management Association). "Chapter 10–Fermi Problems: Guesstimating in Today's Decimal-Point World."

Ross, Joan and Marc Ross. 1986. "Fermi problems, or how to make the most of what you already know." In *Estimation and Mental Computation: 1986 Yearbook*. Edited by H. L. Schoen and M. J. Zweng. National Council of Teachers of Mathematics.

von Baeyer, Hans Christian. 2001. *The Fermi Solution: Essays on Science*. Mineola, NY: Dover Publications. von Baeyer's ruminations on science are interesting and illuminating. His chapter on Fermi problems discusses their history.

30 • Use Forecasts with Care

Armstrong, J. Scott. 20080. *The Forecasting Principles Web Site*, accessible at <http://www.forecastingprinciples.com>.

Bright, James R. 1972. *A Brief Introduction to Technology Forecasting: Concepts and Exercises*. Austin, Texas: The Pemaquid Press. This book is a bit dated in some ways, and its optimism about accurate forecasting is a bit quaint, but it's long on examples, practical advice, and careful thinking about forecasting in the real world.

Ehrlich, Paul R., Anne H. Ehrlich, and John P. Holdren. 1977. *Ecoscience: Population, Resources, Environment*. San Francisco, CA: W.H. Freeman and Company. A beautiful book with a nice writeup of the subtleties of population forecasts, exponential growth, and logistic curves (see Chapter 4). Even though it's more than two decades old, there's still nothing like it.

Institute for Business Forecasting. <http://www.ibf.org>. email: info@ibf.org. 516/504-7576. Great Neck, NY. An institution devoted to disseminating knowledge about business forecasting and planning. If you're interested in forecasting in a business context, this organization is a good place to start.

Morgan, M. Granger, and Max Henrion. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY: Cambridge University Press. One of the best books on uncertainty ever written. A true classic.

Moroney, M. J. 1953. *Facts From Figures*. Harmondsworth, Middlesex, UK: Penguin Books Ltd. This statistics book has an excellent long chapter on correlation versus causality, and a wonderful short chapter on the perils of trend-based forecasting. My favorite quotation from Moroney: "Economic forecasting, like weather forecasting in England, is only valid for the next six hours or so. Beyond that it is sheer guesswork."

Schnaars, Steven. 1989. *Megamistakes: Forecasting and the Myth of Rapid Technological Change*. New York, NY: Free Press. This book gives examples of major forecasting failures. Both lively and insightful, this book is a great read for those interested in pitfalls in forecasting.

Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press. This fascinating book brings empirical data to bear on assessing the accuracy of expert forecasting over two decades, describing what works and what doesn't work in his review of the predictions of about 250 experts over this period.

31 • Hear All Sides

To learn more about Oxford-style debates, go to <http://www.intelligencesquared.com> and <http://www.intelligencesquaredus.org>.

PART V • SHOW YOUR STUFF

General

Williams, Robin. 2003. *The Non-Designer's Design Book: Design and Typographic Principles for the Visual Novice*. 2nd ed. Berkeley, CA: Peachpit Press. This book is a beautiful and useful introduction to design principles. Williams teaches the key concepts of proximity, alignment, repetition, and contrast, all of which are critical to presenting information effectively. She also reviews the use of typefaces.

Hacker, Diana. 2007. *A Writer's Reference*. 6th ed. Boston, MA: Bedford/St. Martins. A superb

sourcebook dealing with the craft of writing, including grammar, spelling, punctuation, references and creating compelling arguments. The associated web site is also a great resource: <http://www.bedfordstmartins.com/Catalog/product/writersreference-seventhedition-hacker>.

32 • Know Your Audience

Heath, Chip, and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. New York, NY: Random House. This book is a terrific summary of what sticks in the mind of readers or listeners. It's surprisingly simple and powerful.

Stebbing, L. Susan. 1961. *Thinking to Some Purpose*. Harmondsworth, Middlesex, UK: Pelican. This book's chapter titled "Difficulties of an audience" gives lessons about critical thinking for effective public speaking before audiences of different types.

33 • Document, Document, Document

Bialik, Carl. 2006. "When a Billion Isn't a Billion." *The Wall Street Journal (online)*. New York, NY. March 30. The Journal's "Numbers Guy" reviews historical and current confusion over the word "billion".

The Chicago Manual of Style: The Essential Guide for Writers, Editors, and Publishers. 2003. 15th ed. Chicago, IL: University of Chicago Press. The classic reference on style and formats, including discussion of spelling, punctuation, grammar, tables, abbreviations, numbers, foreign languages, and other complexities afflicting the modern writer.

Cook, Terry. 1995. "It's 10 O'Clock: Do You Know Where Your Data Are?." *Technology Review*. January. p. 48. An eye-opening article that discusses the real world problems with keeping track of data in the electronic age.

Hacker, Diana. 2006. *A Writer's Reference*. 6th ed. Boston, MA: Bedford/St. Martins. Part of the web site associated with this book has information and examples on research and documentation in the electronic age, with separate guidance for sciences, social sciences, and humanities: <http://www.dianahacker.com/resdoc/>.

Stim, Richard. 2004. *Getting Permission: How to License & Clear Copyrighted Materials, Online and Off*. 2nd ed. Berkeley, CA: Nolo.com. <http://www.nolo.com/>. Practical advice on getting permission to use copyrighted materials.

34 • Let the Tables and Graphs Do the Work

Booth, Wayne C., Gregory G. Colomb and Joseph M. Williams. 1995. *The Craft of Research*. Chicago, IL: The University of Chicago Press. One of the best books on the process of writing research papers, from start to finish.

Strunk Jr., William. 2007. *The Elements of Style*. Claremont, CA: Coyote Canyon Press. The basic *Elements of Style* hasn't changed in decades (it's one of those gems that simply can't be improved upon). Plan on rereading this short but essential book every few years, to refresh your memory and enliven your writing.

The Chicago Manual of Style: The Essential Guide for Writers, Editors, and Publishers. 2003. 15th ed. Chicago, IL: University of Chicago Press. Many writers consider this book to be the last word in rules for writing. It's a comprehensive source, and one that's well worth having on your shelf.

35 • Create Compelling Graphs and Figures

Craver, John S. 1980. *Graph Paper From Your Copier*. Tucson, AZ: HPBooks. If you need a special kind of graph paper in a pinch, this book is just the ticket, with almost 200 different graph and map types, ready for copying and use. It also has a nice 20+ page writeup of hints for making good graphs. It was especially useful before computer graphing became so widespread, but it still can come in handy.

Few, Stephen. 2004. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press. This book is an excellent source of practical graphics advice for business, with reams of examples and clear discussion of key principles.

Jones, Gerald Everett. 2006. *How to Lie with Charts*. 2nd ed. Charleston, SC: BookSurge Publishing. Mainly oriented towards users of computer graphing tools, this book contains dozens of specific recommendations for making effective use of charts and graphs.

Monmonier, Mark, and H. J. de Blij. 1996. *How to Lie with Maps*. 2nd ed. Chicago, IL: University of Chicago Press. This book contains examples of maps that mislead, confuse, advocate, and advertise. If you work with maps a lot, it's a "must read."

Munter, Mary. 2005. *Guide to Managerial Communication*. 7th ed. Englewood Cliffs, NJ: Prentice Hall, Inc. This book describes techniques for effectively using graphs in a management context.

Tufte, Edward R. 1990. *Envisioning Information*. Cheshire, CT: Graphics Press. Tufte says this book is about "*pictures of nouns*," namely graphical representations of places, things, people, and animals. It also deals with "visual strategies for design: color, layering, and interaction effects."

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press. Fourteenth printing. Tufte says this book is about "*pictures of numbers*, how to depict data and enforce statistical accuracy." It is by far the best book from which to learn the subtleties of effective graph design.

Tufte, Edward R. 1997. *Visual Explanation: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press. Tufte says this book is about "*pictures of verbs*, the representation of mechanism and motion, of process and dynamics, of causes and effects, of explanation and narrative." It focuses on the use of graphics to explain events and make explanatory assertions about those events.

Zelazny, Gene. 2001. *Say It With Charts: The Executive's Guide to Visual Communication*. 4th ed. New York, NY: McGrawHill. A fine practical guide to using graphs in the business world.

36 • Create Good Tables

Few, Stephen. 2004. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press. Few gives advice for choosing between tables and graphs, reviews tabular design principles, and presents myriad practical examples.

Jones, Gerald Everett. 2006. *How to Lie with Charts*. 2nd ed. Charleston, SC: BookSurge Publishing. Has a nice section on table design.

Munter, Mary. 2005. *Guide to Managerial Communication*. 7th ed. Englewood Cliffs, NJ: Prentice Hall, Inc. A business-oriented guide to getting your point across.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press. Fourteenth printing. Key design principles and examples of good tables are sprinkled liberally throughout this superb book.

Williams, Robin. 2003. *The Non-Designer's Design Book: Design and Typographic Principles for the Visual Novice*. 2nd ed. Berkeley, CA: Peachpit Press. The design principles in this book (including its introduction to typefaces) are especially pertinent to designing good tables.

37 • Use Numbers Effectively in Oral Presentations

Heath, Chip, and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. New York, NY: Random House. Effective speakers tell stories, and this book teaches how to make them stick in your listeners' minds.

Jones, Gerald Everett. 2006. *How to Lie with Charts*. 2nd ed. Charleston, SC: BookSurge Publishing. Gives specific advice on using tables and graphs in oral presentations.

Osgood, Charles. 1988. *Osgood on Speaking: How to Think on Your Feet Without Falling on Your Face*. New York, NY: William Morrow and Company, Inc. One of the best books on the art of public speaking.

Tufte, Edward R. 1997. *Visual Explanation: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press. In this book, Tufte derives lessons from magicians and illusionists to teach good presentation skills (see pp. 68-70).

38 • Use the Internet

Anderson, Chris. 2006. *The Long Tail: Why the Future of Business is Selling Less of More*. New York, NY: Hyperion. The Internet age has transformed business in many ways, one of which is the movement away from geography and shelf space as primary determinants of the availability of goods. This shift allows business to profitably deliver low-volume products that were simply unavailable to consumers when shelf space was at a premium. "The Long Tail" describes this transformation.

Andersson, Eve, Philip Greenspun, and Andrew Grumet. 2006. *Software Engineering for Internet Applications*. Cambridge, MA: MIT Press. Covers some of the same ground as Greenspun's earlier book, but is more of a textbook and is more up-to-date. For those designing serious Internet applications, it's a must read.

Apple Computer. 1992. *Macintosh Human Interface Guidelines*. Reading, MA: Addison-Wesley Publishing Company. This book gives general information on how computers and humans can most peacefully coexist. The general principles it espouses are useful regardless of which operating system you favor.

Barabasi, Albert-Laszlo. 2003. *Linked: How Everything is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. New York, NY: Plume, a member of Penguin Group (USA). This brilliant book describes the structural similarities of networks that appear on the surface to be quite distinct, including the Internet, the World Wide Web, citation networks, human networks, and the genetic code. Barabasi's brilliant review of the literature describes how "scale-free" networks are transforming the face of business.

Burke, James. 2007. *Connections*. New York, NY: Simon & Schuster. An outstanding book. Among other things, Burke describes the key role of the printing press in fostering later technological and scientific developments.

Greenspun, Philip. 1999. *Phil and Alex's Guide to Web Publishing*. San Francisco, CA: Morgan Kauffman. While somewhat technical in parts, this is absolutely the best book for people interested in making the web useful. It's also a beautiful coffee table book. For the on-line version, see <http://philip.greenspun.com/panda/>.

Norman, Donald A. 1990. *The Design of Everyday Things*. New York, NY: Doubleday/Currency. This book yields important lessons for interface design of just about everything. Don't miss it.

Tapscott, Don, and Anthony D. Williams. 2006. *Wikinomics: How Mass Collaboration Changes Everything*. New York, NY: Portfolio, a member of the Penguin Group USA. This fascinating book documents new collaboration technologies and techniques in a highly readable way.

Williams, Robin. 2003. *The Non-Designer's Design Book : Design and Typographic Principles for the Visual Novice*. 2nd ed. Berkeley, CA: Peachpit Press. This book is a wonderful introduction to design principles, many of which apply directly to web design.

Williams, Robin and John Tollett. 2005. *The Non-Designer's Web Book: An Easy Guide to Creating, Designing, and Posting Your Own Web Site*. 3rd ed. Berkeley, CA: Peachpit Press. A delightful and easy to read book that gives you the basics of web design and implementation, dealing with nitty-gritty details of file formats, typography, colors, effective design strategies, and all sorts of other stuff.

39 • Share and Share Alike

American Library Association. 1999. Task Force on Metadata Summary Report. Committee on Cataloging: Description & Access, Association for Library Collections & Technical Services, American Library Association. CC: DA/TF/Metadata/4. June. <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>. This document gives a precise definition of metadata.

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web" In *Scientific American*. May. pp. 34–43. This reference is the seminal one on the Semantic Web. For more background and details on the latest developments, go to <http://www.w3.org/2001/sw/>, <http://infomesh.net/2001/swintro/>, and <http://en.wikipedia.org/wiki/Semantic_Web>.

Solove, Daniel J. 2007. *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*. New Haven, CT: Yale University Press. The focus of this book is more on personal than professional reputation, but it's still an important one on this topic.

EPILOGUE • SOME PARTING THOUGHTS

Baer, Walter S., Scott Hassell, and Ben Vollaard. 2002. *Electricity Requirements for a Digital Society.* RAND Corporation. MR-1617-DOE, ISBN 0-8330-3279-8. <http://www.rand.org/publications/MR/MR1617/>. This report analyzes current and projects future electricity use by electronic devices. It showed that electricity used by computers and office equipment was only a few percent of the total in 2000 and could not possibly grow to 50% of the total in ten or twenty years.

Holdren, John P. 2002. "Skepticism toward The Skeptical Environmentalist". ScientificAmerican.com, April 15. <http://www.sciam.com/>. Holdren summarized his refutation of Bjorn Lomborg's book in this article.

Huber, Peter, and Mark P. Mills. 1999. "Dig more coal—the PCs are coming." In *Forbes*. May 31. pp. 70–72. *Forbes* started the frenzy over electricity used by the Internet by publishing this article.

Huber, Peter, and Mark P. Mills. 2000. "Got a Computer? More Power to You." *Wall Street Journal*. New York, NY. September 7. p. A26. This article contained the assertion about the networking electricity for a wireless Palm Pilot being equal to that of a residential refrigerator.

Huber, Peter, and Mark P. Mills. 2005. *The Bottomless Well: The Twilight of Fuel, the Virtue of*

Waste, and Why We Will Never Run Out of Energy. New York, NY: Basic Books. This book repeats the claims in the *Forbes* article and adds some others.

Koomey, Jonathan, Kaoru Kawamoto, Bruce Nordman, Mary Ann Piette, and Richard E. Brown. 1999. *Initial comments on 'The Internet Begins with Coal'.* Berkeley, CA: Lawrence Berkeley National Laboratory. LBNL-44698. December 9. (http://enduse.lbl.gov/projects/infotech.html). Our initial response to claims in the *Forbes* article.

Koomey, Jonathan G. 2000. *Rebuttal to Testimony on 'Kyoto and the Internet: The Energy Implications of the Digital Economy'.* Berkeley, CA: Lawrence Berkeley National Laboratory. LBNL-46509. August. <http://enduse.lbl.gov/Projects/InfoTech.html>. This point-by-point rebuttal of testimony of Mark Mills reproduces that testimony and intersperses my responses.

Koomey, Jonathan, Huimin Chong, Woonsien Loh, Bruce Nordman, and Michele Blazek. 2004. "Network electricity use associated with wireless personal digital assistants." *The ASCE Journal of Infrastructure Systems (also LBNL-54105).* vol. 10, no. 3. September. pp. 131–137. This article investigates the claim about network electricity associated with wireless Palm Pilots.

Koomey, Jonathan, Chris Calwell, Skip Laitner, Jane Thornton, Richard E. Brown, Joe Eto, Carrie Webber, and Cathy Cullicott. 2002. "Sorry, wrong number: The use and misuse of numerical facts in analysis and media reporting of energy issues." In *Annual Review of Energy and the Environment 2002.* Edited by R. H. Socolow, D. Anderson and J. Harte. Palo Alto, CA: Annual Reviews, Inc. (also LBNL-50499). pp. 119–158. The electricity used by computers is one of the case studies in this article, which examines how the media treat technical information.

Koomey, Jonathan. 2003. "Sorry, Wrong Number: Separating Fact from Fiction in the Information Age." In *IEEE Spectrum.* June. pp. 11–12. This two-page article summarizes ways to identify and avoid erroneous statistics.

Mills, Mark P. 1999. *The Internet Begins with Coal: A Preliminary Exploration of the Impact of the Internet on Electricity Consumption.* Arlington, VA: The Greening Earth Society. May. Mills documented the claims in the *Forbes* article in this technical report.

Roth, Kurt, Fred Goldstein, and Jonathan Kleinman. 2002. *Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings—Volume I: Energy Consumption Baseline.* Washington, DC: Prepared by Arthur D. Little for the U.S. Department of Energy. A.D. Little Reference no. 72895-00. January. <http://www.eere.energy.gov>. Roth et al. validated our estimate of total electricity used by office equipment (3% of the US total).

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

MORE PRAISE FOR *TURNING NUMBERS INTO KNOWLEDGE*[1]

Turning Numbers Into Knowledge is a godsend to analysts everywhere. It is the single best resource I have encountered on the practice of data analysis in the real world.

Cooper Richey
Trader, Centaurus Energy

This outstanding book teaches the tricks of the analytical trade. There's no better guide to learning how to use numbers to understand the world.

Art Rosenfeld
Commissioner, California Energy Commission
Professor of Physics Emeritus
University of California, Berkeley

I thought I was good at crunching numbers until I read *Turning Numbers Into Knowledge*. It's a great tool for improving your own use of numbers AND for seeing through the smoke screens of others.

Lee Schipper, Ph.D.
World Resources Institute

Turning Numbers into Knowledge teaches tricks that most problem solvers only learn after years on the job. Give yourself an edge: Read this book!

Barbara Barkovich, Ph.D.
Principal
Barkovich and Yap, Inc., Consultants

This gem of a book should be required reading for anyone who analyzes information–and that means everyone!

Professor Richard F. Hirsh
Department of History and
Science & Technology Studies, Virginia Tech

The greatest challenge facing educators, institutions, and businesses is the inculcation of rigor and fact-based analysis into the psyches of future leaders. *Turning Numbers Into Knowledge* is the tool for organizations to accomplish this essential task.

Ephraim Heller
Founder, Therasense, Inc.

Dr. Koomey has produced an absolute home run! A witty, incisive primer on critical analytical thinking. Required reading for business analysts, planners, and strategists–critical insight for players in the Internet economy.

Tod Loofbourrow
President and CEO, Authoria, Inc.

All decision-makers need to read this book. It explains, in clear and useful terms, how to use the ever-growing flow of data in our society. But that's only the start. *Turning Numbers into Knowledge* will help the reader become a better thinker, and it is a rare book that can make that claim.

Hal Harvey
William and Flora Hewlett Foundation

Here at last is the definitive guide for beating information overload and responding to the current anti-science, anti-environment backlash. This remarkable book will empower both professionals and neophytes.

Professor John Harte
Energy and Resources Group, University of California, Berkeley
Author of *Consider a Spherical Cow: A Course in Environmental Problem Solving*

This splendidly clear and concise introduction to the craft should be a foundation of every student's apprenticeship–and for those who missed it, a toolkit for a salutary retrofit later. How much more quickly and pleasantly we would discover truth if everyone followed these simple precepts!

Amory B. Lovins
CEO, Rocky Mountain Institute

Jon Koomey writes books the old fashioned way: by accumulating experiences, anecdotes and examples over a lifetime of hard work, and refining them into a compact ingot of pure gold. The reader is offered a rare gift indeed–the essential elements of dozens of fine books, the collective wisdom of countless scientists and commentators, and a handful of the most inspired comic strips ever to grace the daily newspaper. In an era where so many books, once read, become disposable, Dr. Koomey has created an enduring reference. His focus on the Internet and his commitment to keeping its content fresh through an ongoing electronic dialogue with readers are laudable indeed. Read this book and apply its many lessons not just to school or work, but to *life*!

Chris Calwell
Founder, ECOS Consulting, Inc.

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

NOTES

1. Affiliations listed for informational purposes only. The opinions expressed do not necessarily represent those of the institutions listed.
2. I am indebted to Amory Lovins for this phrase.
3. Tufte, Edward R. 1995. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press. Fourteenth printing. p. 51.
4. By January 2007, the LoC collection totaled more than 130 million items, of which 68% are books and manuscripts <http://www.loc.gov/about/facts.html>. For this calculation I assume that eighty percent of the books and manuscripts are duplicate copies of existing documents, so that our hapless reader would only have to peruse 18 million volumes.
5. An "order of magnitude" is a factor of ten, two orders of magnitude is a factor of one hundred, and so on. This shorthand is commonly used by scientists to summarize large differences in a convenient way.
6. More than half of all scientists who have ever lived are alive today (Stephen Covey, 1992. *Principle-Centered Leadership*. NY, NY: Simon & Schuster. p.91.), and this fact is reflected in ever more rapid rates of scientific and technological progress.
7. Shenk, David. 1997. *Data Smog: Surviving the Information Glut*. New York, NY: HarperCollins Publishers, Inc. The first two paragraphs of this quotation appeared on p. 30 of *Data Smog*. The third paragraph appeared in various electronic versions of Shenk's work, but did not appear in the book itself.
8. In one survey, 80 percent of respondents "felt driven to gather as much information as possible, but over half of them were unable to handle all the information they accumulated." "Survey: New 'dataholics' generation on rise." *San Jose Mercury News*. 8 December 1997. Read on the web at <http://www.sjmercury.com>.
9. Ringer, Robert J. 1974. *Winning Through Intimidation*. New York: Funk & Wagnalls.
10. Hess, Peter, and Julie Siciliano. 1996. *Management: Responsibility for Performance*. New York, NY: McGraw-Hill, Inc. See Chapter 4: Developing goals.
11. Kawasaki, Guy. 1995. *How to Drive Your Competition Crazy: Creating Disruption for Fun and Profit*. New York, NY: Hyperion. p. 46.
12. Kawasaki, Guy. 1995. *How to Drive Your Competition Crazy: Creating Disruption for Fun and Profit*. pp. 40-41.
13. Kawasaki, Guy. 1995. *How to Drive Your Competition Crazy: Creating Disruption for Fun and Profit*. pp. 192-193.
14. Coffee, Peter. 1997. "Don't get sidetracked by trifling differences." *PC Week*. 24 November 1997. p. 32.
15. Hansell, Saul. "In terms of the audience, size matters." NYT May 11, 1998. Read on the web at <http://www.nyt.com>. The Audit Central site, launched in 2000 (but now defunct), standardized the ranking process for sites, so that they could be compared consistently. The prob-

lems with measurement of site traffic continue to this day, as documented in Story, Louise. 2007. "How Many Site Hits? Depends Who's Counting." *The New York Times (online)*. New York, NY. October 22.

16. Reich, Robert. 1997. *Locked in the Cabinet*. New York, NY: Alfred A. Knopf. p. 73.
17. Holton, Gerald. 1973. *Thematic Origins of Scientific Thought: Kepler to Einstein*. Cambridge, MA: Harvard University Press.
18. One famous example is that of Lord Kelvin's estimate of the age of the earth in the mid to late 1800s. Joe D. Burchfield, 1990. *Lord Kelvin and the Age of the Earth*. Chicago, IL: University of Chicago Press. For accounts of dozens of such mistakes, see Youngson, Robert. 1998. *Scientific Blunders: A Brief History of How Wrong Scientists Can Sometimes Be*. New York, NY: Carroll & Graf Publishers, Inc.
19. I made significant effort (and documented this effort extensively) to find the reproduction rights holder for Figure 4.1. If that person or institution contacts me with documentation of ownership, we will negotiate suitable remuneration for use of the figure.
20. The general principles can be derived using either deductive or inductive logic.
21. Edwards, Paul, ed. 1972. *The Encyclopedia of Philosophy*. New York, NY: Macmillan Publishing Co., Inc. & The Free Press. Volume 4, p. 169.
22. Edwards, Paul, ed. 1972. *The Encyclopedia of Philosophy*. Volume 4, p. 169.
23. Sokal, Alan. 1997. "A Plea for Reason, Evidence and Logic." *New Politics* vol. 6, no. 2, pp. 126-129 (Winter 1997).
24. I am indebted to Bill Ahern for prompting me to ask these questions of myself.
25. Covey, Stephen R. 1990. *The 7 Habits of Highly Effective People*. NY, NY: Simon & Schuster. pp. 150-156.
26. Schwartz, Peter. 1996. *The Art of the Long View: Planning for the Future in an Uncertain World*. New York, NY: Doubleday. p.62.
27. In fairness, I made my share of wild assertions in those days.
28. Covey, Stephen R. 1990. *The 7 Habits of Highly Effective People*. NY, NY: Simon & Schuster. pp. 150-156.
29. DeMarco, Tom, and Timothy Lister. 1987. *Peopleware: Productive Projects and Teams*. New York, NY: Dorset House Publishing Company. p. 55.
30. Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. 2d edition. Peterborough, Ontario: Broadview Press. p. 16.
31. Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. pp. 108-110.
32. I am indebted to Professor John Holdren for this phrase.
33. Quoted in Stobough, Robert, Daniel Yergin, I. C. Bupp, Mel Horwitch, Sergio Koreisha, Modesto A. Maidique and Frank Schuller. 1979. *Energy Future*. New York, NY: Ballantine Books.
34. Hughes, William. 1997. *Critical Thinking: An Introduction to the Basic Skills*. pp. 160-161.
35. Meier, Alan. 1987. "Saving the 'Other' Energy in Homes." *Energy Auditor and Retrofitter*. November/December, p. 13.
36. US DOE. 1989. *Residential Energy Consumption Survey (RECS): Housing Characteristics 1987*. EIA, Energy Information Administration, U.S. Department of Energy. DOE/EIA-0314(87). May. <http://www.eia.doe.gov>
37. No matter how rough they may be, Alan's estimates are based on years of experience and professional judgment. They prompt further empirical work that more often than not supports his initial guesses.

38. Calwell, Chris J. 1996. *Halogen Torchieres: Cold Facts and Hot Ceilings*. Boulder, CO: E-Source, Inc. TU-96-10. September.
39. If you subscribe to the *New Republic*, you can read the letter of apology at <http://www.tnr.com/>.
40. For another nice discussion of some common statistical fallacies, check out Barnett, Arnold. 1994. "How numbers are tricking you." *Technology Review*. October. <http://www.technologyreview.com/>
41. 2 Campus Boulevard, Newtown Square, PA 19073, 800/442-0925.
42. I am told that ants use a scented trail to show the way back to the nest. Any scented product (like talcum powder) will confuse them and make them go away.
43. Udell, Jon. 1996. "Web Surveys: Helpful Techniques for Web-based Data Collection and Analysis." *Byte Magazine*. October, p. 133.
44. Crossen, Cynthia. 1994. *Tainted Truth: The Manipulation of Fact in America*. New York, NY: Simon & Schuster, p. 112.
45. Cohen, Bernard L. 1990. *The Nuclear Energy Option: An Alternative for the 90s*. New York, NY: Plenum Press.
46. Reich, Robert. 1997. *Locked in the Cabinet*. New York, NY: Alfred A. Knopf, p.106.
47. Marshall, Jonathan. 1997. "Economists Can't Seem to Agree on Anything." *The SF Chronicle*, August 25, 1997, p. B2.
48. The figure described by this text is contained in "Strategies, Structures, and Processes of Organizational Decisions," which is from *Comparative Studies in Administration*, T. E. Thompson and A. Tuden, (c) 1959. University of Pittsburgh Press.
49. Kunde, Diana. 1998. "Don't Be Bugged." *The SF Chronicle*. Sunday, March 8, 1998, p. J1.
50. Crossen 1994, *Tainted Truth,* p. 106.
51. Crossen 1994, *Tainted Truth,* p. 94.
52. Crossen 1994, *Tainted Truth,* p. 101.
53. Dougan, Michael. 1996. "Ancient quest tests new library." *Sunday SF Examiner*. December 15, 1996, p. C1.
54. For a nice introduction to the concept of a spreadsheet, see Jones, Gerald E. 1995. *How to Lie with Charts*. San Francisco, CA: Sybex, in the section titled "Spreadsheets: The Crash Course," on pages 154-157.
55. CNET online retells the story of this launch as well as other famous computer glitches in "CNET Features—Digital Life—Bugs" (8/24/2000).
56. Sawyer, Kathy. 1999. "Engineers' Lapse Led to Loss of Mars Spacecraft: Lockheed Didn't Tally Metric Units." *The Washington Post*. October 1, 1999, p. A1.
57. Aragon, Lawrence. 1997. "Down with Dirt." *PC Week*. October 27, 1997, p. 83.
58. US raw steel production from 1997 *World Almanac and Book of Facts*, p. 153.
59. U.S. population from *1997 World Almanac and Book of Facts*, p. 382.
60. 1940-1980 GNP from the *Statistical Abstract of the US 1990*, p. 425. 1990 GNP in current $ from *1997 World Almanac and Book of Facts*, p. 133, adjusted to 1982 dollars using the consumer price index from p. 132.
61. The easiest way to calculate this number is to take the ratio of market capitalization per dollar of annual revenues for GE, and divide by the same ratio for Google. This calculation yields 0.16, which equals 1/6.
62. From the perspective of actually *using* the data for calculations (as in this example), downloadable files are almost always preferable to such HTML pages (see Chapter 38: Use the Internet).

63. Ringer, Robert J. 1974. *Winning Through Intimidation*. New York: Funk & Wagnalls.
64. Morgan, M. Granger, and Max Henrion. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY: Cambridge University Press.
65. "Trickle-Down Plan." *The SF Chronicle*. December 5 1996, p. E10.
66. This calculation highlights a startling fact: gasoline is cheaper than bottled water, in spite of its extraordinary energy density. For a nice graph demonstrating this result, see Nadis, Steve and James J. MacKenzie. 1993. *Car Trouble*. World Resources Institute. Washington, DC.
67. Currier, Chet. *The SF Chronicle*. November 25, 1996, p. E3.
68. *The San Francisco Chronicle/Examiner*. Sunday, August 31, 1997. Housing prices were examined on p. E1, and cremation was examined on p. D1.
69. I am indebted to Mark Nowak for this marvelous anecdote.
70. Such intellectual debate is often inhibited between companies because of concerns about proprietary information. These concerns can sometimes be allayed by appropriate use of nondisclosure agreements. Surprisingly, lack of constructive intellectual criticism is the norm even *within* firms, where proprietary concerns are moot.
71. Thanks to Veronica Oliva for her thoughts and comments on this section.
72. "Stamp Price to Rise a Penny to 33 Cents." *The SF Chronicle*. May 12, 1998, p. A1+.
73. Knutson, Lawrence L. 1998. "U.S. to Rain Pennies on Post Office: 33-cent First-class Stamp Takes Effect Next Year." *The Oakland Tribune*. May 12, 1998, p. NEWS-1+.
74. Ward, Sandra. 2007. "Where Is Oil Headed? A Contrarian Says $45." *Barron's*. New York, NY. Number 38, September 17, 2007. p. 56–57.
75. I am indebted to Professor Marcela Bilek for this point. She discovered the importance of it when examining historical trends in Japanese GDP, expressed in dollars. You can investigate for yourself at <http://www.economagic.com>.
76. Herhold, Scott. 1998. "Computer Files Erased at Stanford." *San Jose Mercury News*. April 9, 1998, p. 1A.
77. Tufte, Edward R. 1995. *The Visual Display of Quantitative Information*, p. 51.
78. Tufte, Edward R. 1995. *The Visual Display of Quantitative Information*, p. 170.
79. Vorsatz, Diana, Leslie Shown, Jonathan G. Koomey, Mithra Moezzi, Andrea Denver, and Barbara Atkinson. 1997. *Lighting Market Sourcebook*. Ernest Orlando Lawrence Berkeley National Laboratory. LBNL-39102. December. p. 38. <http://enduse.lbl.gov/Projects/LMS.html>.
80. Tufte, Edward R. 1995. *The Visual Display of Quantitative Information*, p. 191.
81. Tufte, Edward R. 1995. *The Visual Display of Quantitative Information*, p. 191.
82. Vorsatz, Diana, et al. 1997. *Lighting Market Sourcebook*. p. 51.
83. Interlaboratory Working Group. 1997. *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy-Efficient and Low-Carbon Technologies by 2010 and Beyond*. Oak Ridge National Laboratory and Lawrence Berkeley National Laboratory. ORNL/CON-444 and LBNL-40533. September. p. 2.12. <http://enduse.lbl.gov/Projects/5Lab.html>
84. I am particularly indebted to Evan Mills for his detailed comments and suggestions on this chapter.
85. Don Cram at MIT summarizes key literature on event study methods at <http://web.mit.edu/doncram/www/eventstudy.html>.

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

WHY YOU SHOULD READ THIS BOOK

The world keeps getting more complex, but becoming a better problem solver can help you make sense of it all. Mastering the art of problem solving takes more than proficiency with basic calculations: it requires (among other things) understanding how people use information, recognizing the importance of ideology, learning the art of story telling, and acknowledging the important distinction between facts and values. *Turning Numbers into Knowledge* is the first comprehensive guide to these and other essential skills. Full of tools, tricks, and tips for solving problems in the real world, it will prepare you well to make independent judgments about the numerical assertions of others and to generate cogent and compelling analyses of your own.

To order this book or to download key data files, URLs, and sample chapters, go to <http://www.numbersintoknowledge.com>. To find out about related seminars and books, go to <http://www.analyticspress.com>.

ABOUT THE AUTHOR

Credit: Grigorieff Photography

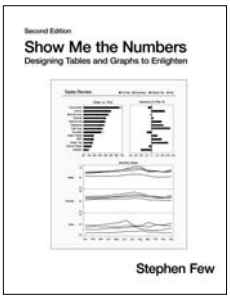
Jonathan Koomey is a Research Fellow at the Steyer-Taylor Center for Energy Policy and Finance at Stanford University. He holds M.S. and Ph.D. degrees from the Energy and Resources Group at the University of California at Berkeley, and an A.B. in History of Science from Harvard University. His academic work, summarized in nine books and more than 150 articles and reports, spans engineering, economics, public policy, and environmental science. He holds a fourth degree black belt in the Japanese martial art of Aikido and enjoys hiking, cooking, and playing classical contrabass in his spare time. For more details go to <http://www.koomey.com>.

ABOUT THE ARTIST

Tom Chen delights in creating sculpture and computer art. You can learn about his sculptures at <http://www.curiousdinosaurs.com>.

ALSO FROM ANALYTICS PRESS

June 2012



Show Me the Numbers: Designing Tables and Graphs to Enlighten, by Stephen Few, is the best selling, simplest, most comprehensive and practical guide to table and graph design available. If you create table and graphs or manage those who do, this book will alleviate countless hours of confusion and frustration.

April 2009

Now You See It, by Stephen Few, does for visual data analysis what Few did for data presentation in *Show Me the Numbers*. It teaches practical techniques that anyone can use—only this time they're for making sense of information, not presenting it.

Copyrighted Materials
Copyright © 2008 Analytics Press Retrieved from www.knovel.com

INDEX

IndexTerms	Links			
A				
Abbey, Edward	211			
Ace, Goodman	83			
Annual Reviews	49	100		
approximations	132	133		
arguments				
counter	61	209		
deductive	25			
inductive	25			
logically strong	59			
missing	61	209		
premises to	27	59	209	
sound	59			
Aristotle	93			
audience, knowing your	147	165	187	194
	211			
authority, questioning	68	209		
B				
Bach, Johann Sebastian	40			
Barrons	157	172	175	
Baruch, Bernard	xv			
beginner's mind	5	14	99	
Berners-Lee, Tim	201			
billion, meaning of	152			
breaking logjams	97	99		
Brynjolfsson, Erik	76			

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
C				
calculations, back-of-the-envelope	41	73	127	129
	188			
Calvin and Hobbes	17	63	210	
Calwell, Chris	74			
Carlin, George	8			
Cartoonbank.com	53	99	154	
Casals, Pablo	52			
Chamberlain, T. C.	61			
Chicago Manual of Style	149	154		
Churchill, Winston	5			
Cialdini, Robert	69	70	91	
Clarke, Arthur C.	112			
Cleaver, Eldridge	32			
Cohen, Bernard L.	87			
cold fusion	24			
Confucius	33			
consistent comparisons	60	105	124	159
	211			
Cook, Rich	128			
copyright	40	149		
Covey, Stephen	53	154		
Creative Commons	207			
critical thinking	13	25	59	69
	89	209		
Crossen, Cynthia	57	83	91	
Cycle of Action	9	31	63	141
D				
data indices	178	181		
data				
archiving	38	151		
backing up	160			
bad	112	187		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
data (*Cont.*)				
cleaning	112	196		
massaging	16			
normalization of	44	114	119	120
	122	123	179	181
public	203			
sharing	190	201		
survey	77	81	91	112
database-backed web sites	200			
databases				
generic	39	42	43	45
	151	191	192	
library	46	101		
reference	40	154		
debates	143			
decision-making				
general	19	42	52	88
	107	151		
using information for	2	11	16	62
	81			
Dilbert	54	133		
documentation				
avoiding bad	123			
creating	13	84	104	149
	194	196	211	
reading the	74			
reasons for	23	36	123	149
	164	177	190	211
sources with good	44	128		
doubling time	139			

E

Economagic	47			
Edison, Thomas A.	160			
Einstein, Albert	28	31	89	

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
Encyclopedia of Philosophy	27	48		
estimates. *See* approximations				
exponential growth	138			
F				
facts vs. values	87	210		
Fed Stats	47			
Few, Stephen	175	179	206	
Fienberg, Stephen E.	86			
figures. *See* graphs				
filing systems	35	38	151	
footnotes, importance of	44	74	84	85
	120	149	152	160
	164	177	196	
Forbes article on IT electricity use	214	219		
forecasting	11	66	109	136
Frequently Asked Questions (FAQs)	41	47	100	
Fuller, Buckminster	200			
G				
General Electric	116	161		
Glass, Stephen	76			
Google	48	116	161	199
	203	205		
graphs				
making	17	43	161	166
reading	84			
Greening Earth Society	216			
Greenspun, Philip	193	194	197	199
	200			
guesses becoming facts	73			
H				
Handbook of Chemistry and Physics	29			
Harte, John	131			

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
Heath, Chip and Dan	5	186		
Heisenberg, Werner	90			
Hoare, C. A. R.	185			
Holdren, John	xiii	51	124	219
Holmes, Sherlock	50	102	104	144
Holton, Gerald	22			
Home Energy Saver	192			
How to Lie with Statistics	42	57	77	172
Huff, Darrell	42	57	77	
Hughes, William	25	57	59	60
	69			
Hyams, Joe	96			
I				
ideology	2	12	18	19
	28	31	63	87
inflation, correcting for	44	156		
information overload	1			
integrity	220			
Intelligence Squared debates	143			
Internet	16	38	46	47
	48	76	100	140
	150	189	211	212
intimidation, avoiding	7	126	163	209
J				
journalists	76	91	137	148
K				
Kawasaki, Guy	13	14	15	
Kay, Alan	110			
Keepin, Will	127			
Kennedy, John F.	143			
Kernighan, Brian W.	165			
Kuhn, Thomas	22	23		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
L				
Land, Edwin	98			
Lee, Bruce	6			
Levine, Mark	153			
librarians, importance of	100			
Library of Congress	1	46	150	
Lieber, Hugh and Lillian	57			
Life in Hell	158			
Lincoln, Abraham	29			
logistic curve	140			
M				
Mattimore, Bryan	95	129	131	
McNealy, Scott	220			
meetings	54			
Meier, Alan	73			
metadata	201			
metric units	49	112	130	153
models				
buried assumptions in	125			
computer	66	109	125	
economic vs. physical science	65			
simple	10	116	125	131
Mosteller, Frederick	86			
N				
National Center for Health Statistics	47			
National Technical Information Service	46			
New Yorker	53	99	154	
newsgroups	31	47		
Nietzsche	79			
Norman, Donald	9	12	15	20
notebooks	40	151		
nuclear power	46	87		

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
O				
oral presentations, using numbers in	147	186		
organization	34	209		
Osgood, Charles	186	188		
Oshima, Tsutomu	148			
P				
Parker, Dorothy	106			
passive voice, avoiding	158			
Pasteur, Louis	36	72		
PDF. *See* Portable Document Format				
Peale, Norman Vincent	59			
Peel, Robert	92			
peer review	22	60	204	
peer reviewed journals	26			
Peopleware	54			
Perot, Ross	83	188		
Pike, Rob	165			
polling results	91			
Polya, G.	98			
Portable Document Format (PDF)	39	150	195	197
	200			
productivity and noise	55			
R				
references, complete	154			
reference database software	40	154		
Reich, Robert	16	87		
reporters. *See* journalists				
Rhymes with Orange	82			
Ringer, Robert	3	7	127	221
	222			
Robbins, John	70			
Roddenberry, Gene	145			

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
Romm, Joe	216			
Roosevelt, Theodore	21			
Rosenfeld, Art	129	147		
Rourke, Robert E.	86			
rule of seventy	139			
S				
scenarios				
creating	18	107	141	211
justifications for creating	128	137	141	
using	12	18	67	107
	132	211		
Schumacher, E. F.	67			
Schwartz, Peter	38	107	108	109
Science Citation Index	25			
scientific proof	25			
scorecard.org	91	192	200	
search engines	38	47	100	191
	192	197	199	201
	206			
Semantic Web	201			
Shenk, David	1			
Sherman's Lagoon	66			
Show Me the Numbers	168	175	179	
Simon, Julian	138			
small multiples	168	196		
Smith, Fred	68			
Software Publishers' Association	76			
Sokal, Alan	27	32		
spreadsheets	42	43	85	102
	113	134	152	163
	190	195	196	201
Stamp, Sir Josiah	75			
Stanford Encyclopedia of Philosophy	48			
Stanley, Bessie A.	64			

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
stateline.org	119	196		
Statistical Abstract of the U.S.	44	46	100	120
	121	156		
Stevenson, Robert Louis	161			
Stoll, Clifford	2	101		
story telling. *See* scenarios				
survey questions	77	81	91	
synthesis	211			
T				
tables				
making	17	43	161	177
reading	84			
Teller, Edward	124			
Thoreau, Henry David	53			
time, valuing your	36	52	95	211
time constraints	12	19	62	
tons, meaning of	152			
troubleshooting calculations	102	111	178	
Tufte, Edward	17	166	167	168
	170	176	177	178
	186	196		
Twain, Mark	68	71	142	213
	222			
U				
U.S. Bureau of the Census	46	47	74	
U.S. Department of Energy	46	73		
U.S. Environmental Protection Agency	47			
U.S. Geological Survey	47			
Ueland, Brenda	7	95		
Uncertainty Principle	62	90		
uncertainty	12	62	107	108
	124	128	132	141
unit analysis	133			

This page has been reformatted by Knovel to provide easier navigation.

IndexTerms	Links			
V				
value choices	10	18	32	63
	87			
van Leeuwenhoek, Antonie	208			
von Baeyer, Hans Christian	129	135		
Vonnegut, Kurt	78			
W				
Walton, Sam	14			
Wikipedia	203	207		
Williams, H. H.	55			
Williams, Robin	178	190	193	197
	199			
World Wide Web	34	38	39	40
	47	82	150	153
	192	201	203	
See also Internet				
Y				
Your Money or Your Life	52			
Z				
Zen in the Martial Arts	96			

This page has been reformatted by Knovel to provide easier navigation.